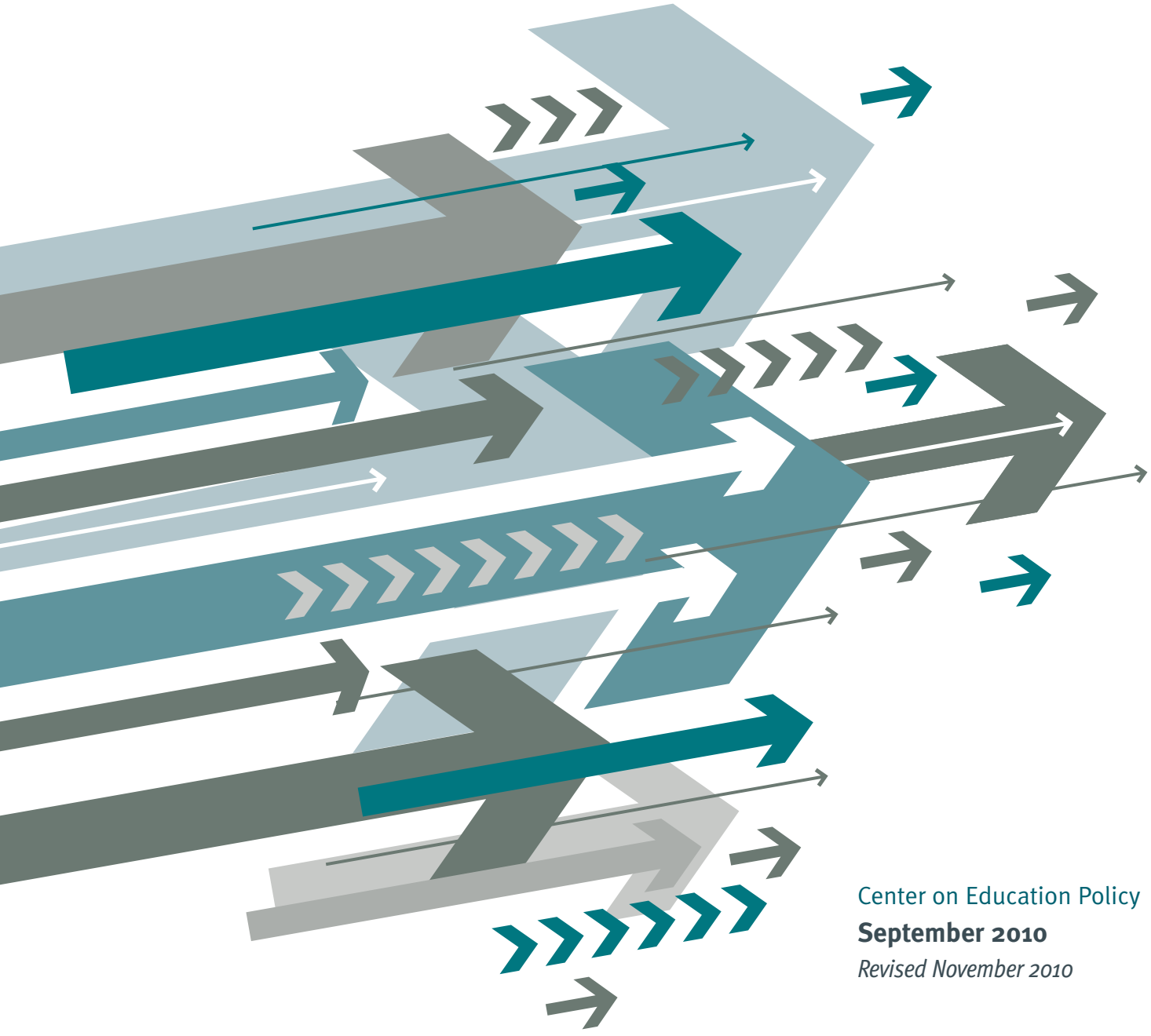
CENTER ON EDUCATION POLICY CEP

State Test Score Trends Through 2008-09, Part 1

Rising Scores on State Tests and NAEP



Center on Education Policy

September 2010

Revised November 2010

Credits and Acknowledgments

This report was written by Naomi Chudowsky and Victor Chudowsky, CEP consultants, who also conducted research for this study. Nancy Kober, CEP consultant, edited the report and assisted with the writing. Sunny Becker, Hilary Campbell, Emily Dickinson, Rebecca Dvorak, and Monica Gribben from the Human Resources Research Organization (HumRRO) provided support in data collection and analysis. Diane Stark Rentner, CEP's director of national programs, oversaw the study project for CEP, communicated with the expert panel, and provided advice and assistance for all aspects of the study. CEP research associate Jennifer McMurrer and CEP interns Malini Roy and Rebecca Johnson assisted with data analysis. Jack Jennings, CEP's president and CEO, provided advice and assistance.

We would like to thank our panel of expert advisors —Laura Hamilton, Eric Hanushek, Frederick Hess, Robert Linn, and W. James Popham—for their invaluable advice. Additionally, we are grateful to the chief state school officers and state assessment personnel for their cooperation in providing information on state testing programs and student achievement data.

Based in Washington, D.C., and founded in January 1995 by Jack Jennings, the Center on Education Policy is a national independent advocate for public education and for more effective public schools. The Center works to help Americans better understand the role of public education in a democracy and the need to improve the academic quality of public schools. We do not represent any special interests. Instead, we help citizens make sense of the conflicting opinions and perceptions about public education and create the conditions that will lead to better public schools.

The Center on Education Policy receives nearly all of its funding from charitable foundations. We are grateful to the Carnegie Corporation, the Smith Richardson Foundation, and the Spencer Foundation for their support of the student achievement study. The George Gund Foundation and the Phi Delta Kappa International Foundation also provide the Center with general support funding that assisted us in this endeavor. The statements made and views expressed are solely the responsibility of the Center.

State Test Score Trends Through 2008-09, Part 1

Rising Scores on State Tests and NAEP

Summary of Main Findings

In recent years, scores on the annual state reading and mathematics tests used for accountability have gone up in most states. These trends in state test scores do not always coincide, however, with trends on the National Assessment of Educational Progress (NAEP), the federally sponsored assessment that is administered periodically to representative samples of students for the nation as a whole and for each state. Consequently, questions arise about which set of assessments is more credible.

This study by the Center on Education Policy (CEP), an independent nonprofit organization, analyzes whether state-level trends in NAEP reading and mathematics results contradict or confirm trends in state test scores. The study focuses on the 23 states with sufficient state test data, meaning states that had comparable data on percentages of students reaching proficiency for 2005 through 2009 for at least one grade/subject combination. For reasons explained later in this report, we compared trends between 2005 and 2009 at grades 4 and 8 in the percentage of students scoring at or above the *proficient* level on state tests and the percentage scoring at or above the basic level on NAEP. We also analyzed achievement on state tests and NAEP using an indicator based on mean, or average, test scores.

We found more agreement between trends on state tests and NAEP than is commonly acknowledged. In general, the majority of states with sufficient data showed gains on both their state test and NAEP. The size of the gains tended to be larger on state tests than on NAEP, however.

Here are the main findings from our study:

- **Since 2005, test scores have increased in most states with sufficient data.** States with test score gains between 2005 and 2009 far outnumbered those with declines on two different assessments (state tests and NAEP) and two different indicators (percentages scoring proficient/basic and mean scores). For example, of the 21 states with sufficient data in grade 8 reading, 20 showed gains in the percentage reaching the proficient level on their state test, and 17 showed gains in the percentage reaching the basic level on NAEP (although the specific states with gains were not always the same for both assessments). Of the 18 states with mean score data on both assessments, 15 showed mean score gains on their state test in grade 8 reading, and 15 exhibited mean score gains on NAEP.

- **Within the same state, trends on NAEP usually moved in the same direction as trends on state tests.** States with positive trends between 2005 and 2009 on their own tests tended to show positive trends on NAEP. This pattern was apparent in percentages proficient/basic and, to an even greater extent, in mean scores. In grade 4 reading, for example, trends on both state tests and NAEP moved in the same direction in 67% of the states with sufficient data using percentages proficient/basic, and in 87% of the states with sufficient data using mean scores. In nearly all cases, trends went up on both assess-

ments. Upward trends on both the state test and NAEP in the same state offer stronger evidence that students are mastering higher levels of knowledge and skills.

- **Gains on state tests tended to be larger in size than gains on NAEP.** This was not always the case, however. In a limited number of states, gains on NAEP were larger than gains on state tests, especially in grade 4 reading.

Background on State Tests and NAEP

State-level data on student achievement in the U.S. come from two primary sources—state tests and the National Assessment of Educational Progress.

Each state has its own testing program, aligned to its own standards for the knowledge and skills that students are expected to learn in key subjects at particular grades. Consistent with the federal No Child Left Behind Act (NCLB), all states must administer their state tests annually to virtually all students in grades 3 through 8 and one high school grade (usually grade 10 or 11). In other important respects, however, including content, difficulty, and format, these tests vary widely from state to state. State tests are considered "high-stakes" assessments because their results are used to hold school districts and schools accountable for students' progress under NCLB and the state's own accountability system. Furthermore, in some states, scores from state tests are used to determine whether students will graduate or be promoted to the next grade.

NAEP, which is overseen by the U.S. Department of Education and is known as "the nation's report card," is designed to track the progress of U.S. students in key subjects at the national and state levels. NAEP encompasses two assessment programs. This report focuses on the main NAEP assessment, which reports national results at grades 4, 8, and 12 and state-by-state results at grades 4 and 8, including trends going as far back as the 1990s. The main NAEP is administered every two years in reading and math and less often in other subjects. The other NAEP assessment program, the long-term trend NAEP, is given every four years in reading and math and reports only national results going back to the 1970s.[1]

NAEP differs from state tests in several important respects:

- **Samples of students versus all students.** NAEP assessments are designed to be administered periodically to representative samples of students in selected schools within each state, rather than annually to virtually all students in a state. Each NAEP participant takes only a portion of the larger assessment instead of the entire test. Consequently, NAEP cannot produce scores for individual students or schools.

- **Different content, format, and administration.** NAEP differs from state tests—to varying degrees, depending on the state—in the content assessed, the test question formats, the rigor of the achievement levels, the testing environment, and other features. In addition, state tests are typically administered by students' own teachers, while NAEP is administered by independent test proctors.

- **Different standards for content.** While state tests are designed to measure how well students have learned the knowledge and skills embodied in each state's academic content standards, NAEP is not deliberately aligned to any state's standards. Rather, NAEP's con-

[1] For a fuller explanation of the differences between the main NAEP and the long-term trend NAEP, see http://nces.ed.gov/nationsreportcard/about/ltt_main_diff.asp.

tent is based on frameworks developed by a National Assessment Governing Board appointed by the U.S. Secretary of Education.

- **Different proficiency definitions.** The term "proficient" often means fundamentally different things on state tests and NAEP. The NAEP definition of proficient is aspirational, signaling where students *should* be in a subject area. Because state tests are used for high-stakes accountability purposes, states are under pressure to set realistic definitions of proficiency that take into account students' current level of achievement. State definitions of proficiency vary; while some are more aspirational than others, most are less ambitious than the NAEP definition. (These differences between the NAEP and state definitions are explained more fully later in this report in box A.)

- **High stakes and low stakes.** NAEP scores are not tied to specific consequences for individual students, teachers, schools, or districts, as state test scores are.

In light of these differences, it is not surprising that the state tests and NAEP sometimes yield different results. When a state test has shown more positive results than NAEP in a particular state, some analysts and policymakers have raised questions about the credibility of the state test scores or dismissed them as overly optimistic. For example, controversy erupted in New York this past year after sizeable gains occurred on the state test while NAEP scale scores remained flat. This situation led some observers to charge that state education officials were making "false claims" about student achievement and were unofficially lowering the number of items students needed to answer correctly to pass or making the tests easier in other unpublicized ways (Ravitch, 2009; Stern, 2010). New York state officials responded by raising the scores needed to pass and making other changes affecting scores from spring 2010 state testing. The percentages proficient dropped dramatically, leading to confusion, surprise, or anger among parents, students, and educators (New York State Department of Education, 2010; New York Daily News, 2010; Medina, 2010).

The New York controversy is part of a larger ongoing debate among policymakers and researchers about the extent to which gains in state test scores reflect real increases in learning. By "real" increases in learning, we mean that students have acquired knowledge and skills tied to valued educational goals, not just the specific content measured by a particular test. Serious consequences are attached to poor results on state tests, such as bad publicity, replacement of teachers and principals, major changes in school governance and management, and even failure of students to graduate in some states. In this high-stakes testing environment, teachers and administrators have strong incentives to raise test scores and may choose to do so by the easiest means possible. Because tests are able to cover just a sample of the content included in a particular subject, teachers may have a tendency to focus instruction only on the material that is likely to be tested at the expense of other material in the same subject or different content and educational goals in other subjects. They may directly coach students on test-taking skills and the content likely to show up on a high-stakes test or may even engage in outright cheating. These practices can lead to exaggerated gains on the state test, which researchers refer to as "score inflation."

If gains on a particular test reflect real gains in learning, researchers expect to see some degree of "generalization" across assessments in the same subject (Koretz, 2005). This means that students have mastered enough of the knowledge and skills in a particular domain, such as grade 4 math, that they can perform better not just on a high-stakes test but on other tests and non-test indicators of the same domain. If high scores do not generalize to other measures of achievement, that is one clue that students may be learning only the narrow part of the domain that is included on a particular test.

NAEP is often viewed as a kind of "audit" of state tests because it offers a critical perspective on student achievement that is independent of state tests. Policymakers should be aware, however, of NAEP's limitations in this role. First, students may not be motivated to perform their best on NAEP, since NAEP does not produce individual scores, is not taken by all students, and is not tied to specific consequences. The administration of NAEP by outside proctors could also affect students' motivation or anxiety in unknown ways. Similarly, teachers and administrators may be less motivated to prepare students for NAEP than for the higher-stakes state tests.

Second, NAEP may not assess what students are actually taught in the classroom because it is not aligned to any state's content standards, whereas state tests are aligned to each state's standards to varying degrees. Furthermore, teachers are unlikely to tailor their instruction to NAEP assessments due to the difficulty of knowing which material might be tested. For state tests, however, teachers often try to mesh their instruction with the likely content of the test.

For these reasons, NAEP results should not be treated as if they override or invalidate state test results. Rather, NAEP offers an additional source of information that can be used in conjunction with state test data to gain a fuller picture of student achievement in a specific state. Indeed, comparisons of trends on state tests and NAEP are informative precisely because NAEP is a low-stakes measure of achievement without all of the external pressures and incentives attached. We conducted this study comparing state tests and NAEP in the spirit of recognizing the value and limitations of both types of assessments.

Purpose of This Study and Approach Used

Some past studies have shown little relationship between gains in state test scores and NAEP results over various time spans (e.g., Fuller et al., 2006; Jacob, 2007; Koretz, 2005). Our previous study of this kind, which looked at state and NAEP trend data from as early as 2002 through 2007,[2] found that while gains occurred on both state tests and NAEP, gains in state test scores were larger in size than gains on NAEP (CEP, 2008).

This study updates our earlier study by including state test data from school year 2008-09, the most recent year available at the time of our data collection, and from the 2009 administration of NAEP. The addition of two more years of data has created longer trend lines in most states and enabled us to see whether the trends identified in our earlier study of state tests and NAEP have held up. This study looks at three key questions: Do NAEP trends contradict or confirm state trends? Do gains in state test scores also show up as gains on NAEP in the same state? Does a large increase in state test scores mean a large increase on NAEP?

Several issues can complicate comparisons of state and NAEP results and cause confusion among policymakers, the media, and the public. To address these sometimes complex issues, we relied on advice from a panel of educational testing and policy experts who have assisted us with all of our student achievement studies.[3] With their help, we arrived at the following approach to compare state test and NAEP trends:

[2] The specific span of years analyzed in our previous study varied by state because many states lacked comparable state test data going back to 2002.

[3] Members of the expert panel include Laura Hamilton, senior behavioral scientist, RAND Corporation; Eric Hanushek, senior fellow, Hoover Institution; Frederick Hess, director of education policy studies, American Enterprise Institute; Robert L. Linn, professor emeritus, University of Colorado; and W. James Popham, professor emeritus, University of California, Los Angeles.

- **State-level results from main NAEP.** Since states were the unit of analysis for this study, we compared state-by-state results on the main NAEP with results from state tests in each state with a continuous trend line from 2005 through 2009. As noted above, the long-term trend NAEP cannot be used for state-level comparisons because it does not report state-level results.

- **Subjects and grades.** For both state tests and NAEP, we examined trends at grades 4 and 8 in reading and math, the subjects tested for NCLB accountability. (Utah uses an end-of-course test of pre-algebra as its grade 8 test, which students take after they have completed the appropriate course.) High school results were not analyzed because NAEP data at the high school level are not broken down by state and because NAEP is given in grade 12, whereas most state tests are administered in grade 10 or 11.

- **Years analyzed and number of states included.** Our primary analyses compared trends on state tests and NAEP from 2005 through 2009, the same time period for both tests. Twenty-three states had continuous state test data for that period and could be included in the analyses. (All states have NAEP data for 2005, 2007, and 2009). The other states had "breaks" in their test data because they had introduced new tests or changed their cut scores for proficient performance; with these types of breaks, year-to-year comparisons are not valid. As a secondary analysis, we also examined trends from 2007 through 2009 because we had almost twice as many states (43 states) with sufficient data for this period. However, we placed more weight on the 2005–2009 findings because, in general, longer trend lines tend to be more reliable for determining achievement trends (Kane & Staiger, 2002; Linn & Haug, 2002). State test scores can fluctuate from year to year for reasons unrelated to teaching and learning, such as shifts in the population of students being tested each year—for instance, if a state experiences an influx of immigrants or a drop in employment. In addition, one-time factors such as a teacher strike or flu epidemic can cause fluctuations (Linn & Haug, 2002). A longer trend line makes it more possible to see cumulative effects across years rather than short-term fluctuations. The reason we did not go back further, to the 2003 NAEP administration for instance, is because a much smaller number of states had continuous trend lines for this period.

- **Comparisons of state percentages proficient with NAEP percentages basic.** Both state tests and NAEP report their results in terms of various achievement levels, such as basic, proficient, and advanced, but the definitions, names, and number of levels vary among states and between state tests and NAEP. As explained in **box A**, the term "proficient" represents two fundamentally different concepts for NAEP and state tests. For NAEP, "proficient" represents an aspirational goal for what student should know and be able to do, while on most state tests, it describes the level of student performance that is good enough to be regarded as acceptable for a particular grade level. As explained in box A, it is most appropriate to compare percentages of students scoring at or above the proficient level on state tests with percentages scoring at or above the basic level on NAEP.

- **Mean score comparisons.** We also compared trends in mean (average) scale scores on state tests with those on NAEP. (A mean score is the average of a group of test scores expressed on a common scale for a particular state's test; it is calculated by adding the scores and dividing the sum by the number of scores.) All tests report results on a numerical scale, but they use different scales, such as 1–100 or 1–500. Unlike percentages proficient, mean scale scores do not depend on where cut scores are set. Mean scores also pick up improvements along the entire scoring scale, not just at the proficient or basic levels. For some analyses, we used mean scores to compute a statistic called

effect size, which allows one to compare changes between two different tests with different scoring scales.[4]

- **Use of average yearly gains or declines.** Our analyses focused on trends over time. In particular, we determined whether results on a state test and on NAEP had improved, declined, or showed no change over a certain period, regardless of whether scores started out or ended up higher on one test or the other. For each state with sufficient data, we compared average yearly changes in state test results with average yearly changes in NAEP results. To calculate these averages, we divided the overall change in the percentage proficient/basic or mean scale score by the number of years covered by the trend.

- **No tests of statistical significance for changes.** NAEP tests a sample of students in a sample of schools in each state and computes statistical estimates of student performance to generalize results from this sample to the state's entire student population. The NAEP program is understandably careful to report the degree of confidence that data users should have in these sample-based estimates and highlights shifts in performance only when they are statistically significant. State tests, by contrast, are administered to virtually all students in a particular grade; checks for statistical significance are not necessary or appropriate because state test results already represent the entire student population and do not have to be extrapolated from a sample. In this study, we interpreted trends on NAEP in much the same way as trends on state tests, counting an increase or decrease of any size as a gain or decline. We did not constrain comparisons by limiting NAEP data to statistically significant changes. To do otherwise would mean judging state tests and NAEP by different rules. However, because we are counting even small changes as increases or decreases, it is possible that some of these merely reflect random fluctuations in some states.

Box A. Why compare the state proficient level with the NAEP basic level?

Both NAEP and state testing programs report results using multiple levels of student achievement. NAEP has defined three achievement levels—basic, proficient, and advanced. States are required by NCLB to establish a minimum of three achievement levels on their state tests—often called basic, proficient, and advanced but sometimes labeled differently. In most states, the percentages of students reaching the proficient level in math and reading are the main indicators used to determine progress for federal accountability purposes. However, the proficient level on most state tests is not readily comparable to the proficient level on NAEP. As explained below, it is more appropriate to compare the percentage scoring at or above the proficient level on state tests with the percentage scoring at or above the basic level on NAEP.

Although the label is similar, the term "proficient" means fundamentally different things on state tests and on NAEP. The NAEP definition of proficient is aspirational, signaling where the National Assessment Governing Board (NAGB) believes students *should be* and embodying the knowledge and skills that NAGB believes should be included in a well-designed curriculum for that subject area. To reach the NAEP proficient level, students must demonstrate "solid academic performance" and "competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter." To reach the NAEP "basic" level, students must demonstrate "partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade" (National Assessment Governing Board, n.d.).

(continued)

[4] An effect size is a statistical tool that conveys the amount of difference between test results using a common unit of measurement which does not depend on the scoring scale for a particular test. We computed an effect size statistic called Cohen's D. This is done by subtracting the year 1 mean test score from the year 2 mean test score and dividing by the average standard deviation of the two years. (The standard deviation is a measure of how much test scores tend to deviate from the mean—in other words, how spread out or bunched together scores are.) Where there has been no change, the effect size is 0. An effect size of +1 indicates a shift upward of one standard deviation from the previous year's mean test score. In practice, effect sizes tend to be much smaller than 1 for year-to-year changes. To determine trends over multiple years, we calculated the cumulative change in effect size after calculating the year-to-year changes in effect size.

State definitions of proficient, by contrast, are tied to the state's content standards and vary considerably across states, as does the content of these tests. Nevertheless, because state tests are used for high-stakes accountability purposes, all states are under pressure to set realistic definitions of proficiency that take into account students' current level of achievement as well as public perceptions. If tests and cut scores for proficiency are too easy, that may result in very high percentages proficient that are not seen as credible by the public, policy analysts, or researchers. (This in fact has happened in some states.) If the tests and cut scores are too difficult, then massive numbers of students may fail to reach the proficiency threshold, which could be unpalatable. In the majority of states, percentages proficient are above 70%.

This difference between the aspirational goals of NAEP and the more realistic goals of many state tests has led to considerable confusion. A 2009 "mapping" study by the National Center for Education Statistics (NCES), which administers NAEP, placed states' standards for proficiency onto the NAEP scoring scales. The mapping study provided evidence that in most states, cut scores for proficient performance on state tests were less ambitious than the NAEP proficient level and often were closer to—or sometimes below—the NAEP basic level (Bandeira de Mello, Blankenship, & McLaughlin, 2009).

Our data support the conclusions of NCES. The table below gives a snapshot of the 2009 percentages proficient in grade 4 and grade 8 reading and math and compares them with the percentages basic and proficient on NAEP. For each grade/subject combination, the table shows the median[5] percentages of students reaching these various levels, along with the lowest percentage in any state (the minimum) and highest percentage (the maximum). In each grade/subject combination, the median percentage proficient on state tests is much closer to the percentage for NAEP basic than NAEP proficient.[6]

Percentages of students reaching the proficient level on state tests and the basic and proficient levels on NAEP, 2009

READING	Grade 4			Grade 8		
	State proficient	NAEP basic	NAEP proficient	State proficient	NAEP basic	NAEP proficient
Median	74%	69%	33%	71%	77%	32%
Minimum	45%	44%	17%	45%	51%	14%
Maximum	95%	80%	47%	95%	86%	43%
MATH	**Grade 4**			**Grade 8**		
	State proficient	NAEP basic	NAEP proficient	State proficient	NAEP basic	NAEP proficient
Median	74%	84%	40%	66%	75%	35%
Minimum	42%	56%	17%	39%	40%	11%
Maximum	96%	92%	57%	92%	86%	52%

Table reads: Across states in 2009, the median percentage of students performing at or above the proficient level on their state's test was 74% in grade 4 state reading. On the NAEP grade 4 reading test, the median percentage of students performing at or above the NAEP basic level was 69%, and the median performing at or above the NAEP proficient level was 33%.

[5] The median is the middle number in a list of numbers ordered by value, so that half of the numbers in the list are greater in value than the median and half are less. As used in this report, the median percentage proficient or basic for a specific subject and grade (such as grade 8 math) represents the midpoint across all of the states with sufficient data; half of these states had percentages above the median and half had percentages below.

[6] The median percentage proficient for state tests is based on many different state tests of varying difficulty, whereas the median percentages basic and proficient on NAEP are based on the same test administered across all states.

The tables in the body of this report display the total number of states included in each analysis and the numbers of states showing various trends. For readers interested in seeing which specific states demonstrated which trends, **appendix 1** includes more detailed versions with state names for most of the tables in this report. **Appendix 2** contains state-by-state tables showing the percentages proficient on state tests and the percentages basic on NAEP for the 23 states included in the 2005–2009 proficient/basic analyses in this report.

Direction of Trends on State Tests and NAEP

Test scores have increased in most of the states analyzed for this study. Between 2005 and 2009, states with gains far outnumbered those with declines on state tests and NAEP and on two different indicators of achievement, percentages scoring proficient/basic and mean scores.

Our previous studies of student achievement have documented increases in state reading and math test scores since 2002 in a large majority of states at the elementary, middle, and high school levels (CEP, 2007; 2008; 2009). This study seeks to shed more light on achievement trends by examining the consistency of state test and NAEP trends since 2005 in those states with comparable state test data for 2005 through 2009.

Altogether, 23 states have comparable percentage proficient data on their state tests for this period; these include Alabama, Alaska, Arizona, Arkansas, California, Colorado, Florida, Iowa, Louisiana, Maryland, Massachusetts, Montana, Nebraska, Nevada, New Mexico, North Dakota, Ohio, Pennsylvania, Tennessee, Texas, Utah, Washington, and Wisconsin. (All states have NAEP data.) The reason this number is not higher is because many states have adopted changes in their tests or their proficiency cut scores that make it inappropriate to compare results from previous years' tests. Furthermore, a few of these 23 states lacked sufficient data for one or more grade/subject combinations; in grade 4 math, for example, 19 states had sufficient data. Two additional states, Delaware and Oregon, lacked comparable percentage proficient data on their state test but did have mean scores for at least one grade/subject; these states are included in the mean score analyses in this report.

Between 2005 and 2009, most of the states with sufficient data made gains in the percentage of students scoring at the proficient level on state tests and the percentage scoring at the basic level on NAEP, as shown in **table 1**. On both assessments, states with gains far outnumbered those with declines. In grade 4 reading, for example, 16 of 21 states, or 76% of the states with sufficient data, showed gains on the state test. The same number showed gains on NAEP, although these were not necessarily the same 16 states. (As mentioned above, detailed versions with state names of most of the tables in the report can be found in appendix 1.)

Although the percentage of students scoring at the proficient level is important for accountability and public reporting purposes, it only captures student performance at a certain point on the achievement spectrum. By contrast, mean scores capture the performance at all levels, high and low. Percentages proficient can go up without an increase in mean test scores—for example, when some students improve enough to cross the proficiency threshold but students at the higher or lower ends of the achievement spectrum do worse. Therefore, as a check on the percentage proficient/basic results, we also calculated gains and declines using mean test scores.

Table 1. Number (and percentage) of states with gains and declines on state tests and NAEP from 2005 to 2009

Subject, grade, trend	State proficient trend		NAEP basic trend	
Grade 4 reading				
# of states with sufficient data	21		21	
# of states with gains	16	*(76%)*	16	*(76%)*
# of states with declines	3	*(14%)*	3	*(14%)*
# of states with no change	2	*(10%)*	2	*(10%)*
Grade 8 reading				
# of states with sufficient data	21		21	
# of states with gains	20	*(95%)*	17	*(81%)*
# of states with declines	1	*(5%)*	1	*(5%)*
# of states with no change	0	*(0%)*	3	*(14%)*
Grade 4 math				
# of states with sufficient data	19		19	
# of states with gains	18	*(95%)*	15	*(79%)*
# of states with declines	1	*(5%)*	2	*(11%)*
# of states with no change	0	*(0%)*	2	*(11%)*
Grade 8 math				
# of states with sufficient data	21		21	
# of states with gains	20	*(95%)*	20	*(95%)*
# of states with declines	1	*(5%)*	0	*(0%)*
# of states with no change	0	*(0%)*	1	*(5%)*

Table reads: Of the 21 states with sufficient state test data in grade 8 reading, 20 states (95%) showed gains between 2005 and 2009 in the percentage of students reaching the proficient level on state tests. Seventeen of these 21 states (81%) showed gains during this period in the percentage of students reaching the basic level on NAEP.

A smaller pool of 19 states provided mean score data in one or more grade/subject combinations for 2005 through 2009; these include Alabama, Arizona, Arkansas, California, Colorado, Delaware, Florida, Iowa, Louisiana, Montana, Nevada, New Mexico, North Dakota, Oregon, Pennsylvania, Tennessee, Texas, Utah, and Washington. Not all of these states had mean score data for all subjects and grades, however. The results of the mean score analysis confirmed the general pattern of the percentages proficient/basic analysis. The large majority of states showed gains in mean scores on state tests, and the same was true on NAEP. For instance, of the 18 states with mean score data in grade 8 reading, 16 showed gains and 2 showed declines on their state test; 15 states had mean score gains and 3 states had declines on NAEP. Of the 17 states with mean score data in grade 8 math, 16 reported gains and 1 reported a decline on their state test; all 17 states showed mean score gains on NAEP.

There was greater divergence between state tests and NAEP over the shorter time period from 2007 to 2009 than there was from 2005 to 2009. Forty-five states had percentage proficient/basic data for at least some subjects and grades for the 2007–2009 time frame; these included all states and the District of Columbia except Indiana, Mississippi, New Jersey, Oklahoma, South Carolina, and West Virginia. More states showed declines between 2007 and 2009 on NAEP than on state tests. For instance, in grade 4 reading, 9 of the 43 states with data showed declines on their state tests, while 21 showed declines on NAEP. In grade 4 math, 8 out of 43 states with data had declines on their state tests and 14 had declines on NAEP.

For the 2007–2009 period, 40 states and D.C. had mean score data; the exceptions were Indiana, Maryland, Massachusetts, Mississippi, Nebraska, New Jersey, Ohio, Oklahoma, South Carolina, Virginia, and West Virginia. Again, there was greater divergence between state tests and NAEP during this shorter period than during the 2005–2009 span. As gauged by mean scores, 8 of the 38 states with grade 4 reading data had declines on state tests, but on NAEP 21 had declines. In grade 4 math, 8 states showed declines on state tests, while 18 showed declines on NAEP. As noted above, however, we assign more weight to the longer trend lines because they are generally better for determining real achievement trends.

Within the same state, trends on NAEP usually moved in the same direction as trends on state tests between 2005 and 2009. States with positive trends on their own tests tended to show a positive trend on NAEP—a pattern that was even more apparent for mean scores than for percentages proficient/basic.

The analysis summarized in table 1 revealed the number of states with gains or declines on state tests and NAEP but does not highlight whether the two measures were in sync with one another in the same state. For example, 16 states had increases on state tests, and 16 had increases on NAEP, but these were not always the same states. To see how much overlap occurred in the same state, we compared the direction of state trends and NAEP trends between 2005 and 2009 for each of the states with sufficient proficient/basic data.

For each subject/grade combination, we grouped states as follows:

1. **Trends agree:** Both state and NAEP trends moved in the same direction.
   a. Both up: Both assessments showed increases.
   b. Both down: Both assessments showed declines.

2. **Trends disagree:** States showed a gain on one assessment but a decline on the other.

3. **One flat:** One assessment showed no change, while the other showed either a gain or decline.[7]

The results are depicted in **table 2**. In general, NAEP trends moved in the same direction as state test score trends, although the extent of agreement varied by grade/subject combination. In grade 4 reading, for example, trends on the two assessments moved in the same direction 67% of the time, while in grade 8 math they agreed 90% of the time. In all but one state, the agreement occurred because trends on both assessments went up.

[7] We also looked for states in which trends on both state tests and NAEP showed no change, but found no such instances.

Table 2. Extent of agreement between state tests and NAEP in percentage proficient/basic trends, 2005 to 2009

READING	Grade 4	Grade 8
Number of states with sufficient data	21	21
Number of states where trends agree	14	16
Both up	13	16
Both down	1	0
Trends disagree	3	2
One flat	4	3
Percentage of states in agreement	67%	76%
MATHEMATICS	**Grade 4**	**Grade 8**
Number of states with sufficient data	19	21
Number of states where trends agree	15	19
Both up	15	19
Both down	0	0
Trends disagree	2	1
One flat	2	1
Percentage of states in agreement	79%	90%

Table reads: Of the 21 states with sufficient data in grade 4 reading, trends in the percentages of students reaching the proficient level on state tests and the basic level on NAEP moved in the same direction between 2005 and 2009 in 14 states. Thirteen of these states made gains on both the state test and NAEP, while one state showed a decline on both assessments. Altogether, trends on state tests and NAEP moved in the same direction in 67% of the states with sufficient data in grade 4 reading.

Our analysis of mean scores revealed even greater agreement in the direction of trends on the two assessments. As shown in **table 3**, trends on NAEP agreed with trends on state tests in 87% of the states with sufficient data in grade 4 reading, 83% in grade 8 reading, 79% in grade 4 math, and 94% in grade 8 math. In nearly all the cases of agreement, both state test and NAEP trends went up. This greater level of agreement probably occurs because mean scores represent the middle of the score distribution and are influenced by all test scores, whereas percentages proficient depend on where the state has set its proficiency cut score; if the proficient (or basic) cut score is farther from the middle of the distribution (the mean), then percentages proficient (or basic) may be more subject to the kinds of random fluctuations that testing experts refer to as "measurement error."

When trends on the state test and NAEP have both moved upward in the same state, this offers a stronger base of evidence that students have actually mastered higher levels of knowledge and skills.

Table 3. Extent of agreement between state tests and NAEP in mean score trends, 2005 to 2009

READING	Grade 4	Grade 8
Number of states with state and NAEP data	15	18
Number of states where trends agree	13	15
Both up	11	14
Both down	2	1
Trends disagree	2	3
One flat	0	0
Percentage of states in agreement	87%	83%
MATHEMATICS	**Grade 4**	**Grade 8**
Number of states with state and NAEP data	14	17
Number of states where trends agree	11	16
Both up	11	16
Both down	0	0
Trends disagree	3	1
One flat	0	0
Percentage of states in agreement	79%	94%

Table reads: Of the 15 states with sufficient data in grade 4 reading, trends in mean scores from 2005 through 2009 moved in the same direction on state tests and NAEP in 13 states. Eleven of these states made gains on both the state test and NAEP, while two states showed a decline on both assessments. Altogether, trends on state tests and NAEP moved in the same direction in 87% of the states with sufficient data in grade 4 reading.

We found notably less agreement in the direction of trends on state tests and NAEP for the period from 2007 through 2009 than we did for 2005 through 2009, as displayed in **table 4**. According to percentages proficient/basic, trends on the two assessments moved in the same direction in just 35% of the states with sufficient data in grade 4 reading but ranged as high as 67% of these states in grade 8 math. Mean score trends for 2007–2009 were in sync more often than percentage proficient/basic trends; the share of states with sufficient data that showed state test and NAEP mean score trends moving in the same direction ranged from 55% in grade 4 math to 82% in grade 8 math. The caveat noted above also applies to these findings: trends are less reliable over the shorter span of 2007–2009, so we give more weight to the 2005–2009 results.

Table 4. Extent of agreement on state tests and NAEP, 2007 to 2009

READING	Grade 4	Grade 8
Percentage proficient/ basic		
# of states with state and NAEP data	43	43
# in which state & NAEP trends agree	21	23
% in which state & NAEP trends agree	49%	53%
Mean scores		
# of states with state and NAEP data	38	38
# in which state & NAEP trends agree	22	25
% in which state & NAEP trends agree	58%	66%
MATHEMATICS	**Grade 4**	**Grade 8**
Percentage proficient/ basic		
# of states with state and NAEP data	43	43
# in which state & NAEP trends agree	15	29
% in which state & NAEP trends agree	35%	67%
Mean scores		
# of states with state and NAEP data	38	38
# in which state & NAEP trends agree	21	31
% in which state & NAEP trends agree	55%	82%

Table reads: Of the 43 states with sufficient data in grade 4 reading for 2007 through 2009, trends in the percentage proficient on state tests and in the percentage basic on NAEP moved in the same direction during this period in 21 states, or 49% of these states.

Size of Gains

Gains on state tests tended to be larger in size than gains on NAEP, although NAEP gains were larger than state test gains in some states.

Our previous study comparing trends in scores on state tests and NAEP found that states with increases on both assessments tended to have larger gains on the state tests than on NAEP (CEP, 2008). To see whether this was still the case, we looked at average yearly gains in percentages proficient/basic and in effect sizes, a statistic based on mean scores,[8] for both state tests and NAEP.

As shown in **table 5**, the majority of states with gains on at least one assessment between 2005 and 2009 had larger gains on state tests than on NAEP, although in some cases the differences in the size of gains between the two assessments were small. (States with an increase on one assessment and a decrease on the other were considered to have larger gains on the assessment with the increase.)

States with greater gains on state tests outnumbered those with greater gains on NAEP for all grade/subject combinations, whether we looked at the larger number of states with per-

[8] Effect sizes are explained in more detail in the section above on the approach used for this study.

Table 5. Number (and percentages) of states in which gains from 2005 through 2009 were larger on state tests or NAEP

READING	Grade 4	Grade 8
Proficient/basic trend		
# of states with gains on one or both assessments	19*	21*
State gain > NAEP gain	12 *(63%)*	16 *(76%)*
NAEP gain > state gain	6 *(32%)*	4 *(19%)*
Mean score (effect size) trend		
# of states with gains on one or both assessments	13*	17*
State gain > NAEP gain	8 *(62%)*	14 *(82%)*
NAEP gain > state gain	4 *(31%)*	2 *(12%)*
MATHEMATICS	**Grade 4**	**Grade 8**
Proficient/basic trend		
# of states with gains on one or both assessments	18	16*
State gain > NAEP gain	14 *(78%)*	13 *(81%)*
NAEP gain > state gain	4 *(22%)*	2 *(13%)*
Mean score (effect size) trend	**Grade 4**	**Grade 8**
# of states with gains on one or both assessments	14*	17*
State gain > NAEP gain	9 *(64%)*	12 *(71%)*
NAEP gain > state gain	4 *(29%)*	3 *(18%)*

Table reads: Of the 19 states with gains on at least one assessment (the state test and/or NAEP) in the percentages of students scoring proficient/basic in grade 4 reading, the gain was larger on the state test than on NAEP in 12 states and was larger on NAEP than on the state test in 6 states.

*The numbers below do not add up to the total number of states with gains because some states had the same size gains on the state test and NAEP.

centage proficient/basic data or the smaller pool of states with effect size (mean score) data. In grade 8 reading, for example, the percentage proficient gain on the state test was larger than the percentage basic gain on NAEP in 16 of 21 states but was smaller than the NAEP gain in 4 states. For that same grade and subject, the gain in effect size was larger on the state test in 14 of 17 states but larger on NAEP in 2 states. (In the remaining states, the gains were the same size on the state test and NAEP.)

Across all grades and subjects, gains on state tests were larger than gains on NAEP in 74% of the instances we analyzed using the percentage proficient/basic indicator, while NAEP gains exceeded state test score gains in 23% of these instances. (By "instance" we mean a trend for a particular subject and grade in one state.) A similar pattern was also apparent across all grades and subjects using effect sizes: gains in effect size were greater on state tests than on NAEP in 72% of the instances analyzed, while the reverse was true in 22% of instances. In a small percentage of instances, the gains on both tests were the same.

Some differences emerged by grade level and subject, as shown in table 5. In grade 4 reading, a notable minority (32%) of the states with sufficient data showed larger gains on

NAEP than on state tests. For most other grade/subject combinations, NAEP gains were larger than state test gains in all but a handful of states.

We were interested in knowing more about the states that showed larger gains on NAEP than on their state tests because this finding is inconsistent with other evidence suggesting that state test scores are sometimes inflated. We looked for any similarities among the states that had consistently smaller gains on state tests than on NAEP across subjects and grades. Although no state had smaller gains on its own test than on NAEP for all grade/subject combinations, four states exhibited this pattern in two or more of the four grade/subject combinations; these include Alaska, Colorado, New Mexico, and Tennessee. We hypothesized that these states might be seeing smaller gains on their own tests because their tests were easy or they had low cut scores. If this were the case, their percentages proficient would already be high, leaving little room for improvement (often referred to as the "ceiling effect"). This indeed was the case in Colorado and Tennessee, in which more than 80% or 90% of students scored proficient on state tests, depending on the grade and subject. But Alaska and New Mexico did not have very high percentages proficient. A more detailed study of their testing programs would be needed to explore why they produced smaller gains on state tests.

We also did the same analysis of the size of gains for the period from 2007 through 2009. The results were similar; in most states, state test scores gains were larger than NAEP gains.

In addition to comparing the size of gains on state tests and NAEP for individual states, we also compared the median changes in scores on the two assessments between 2005 and 2009 across *all* of the states with sufficient data. The median is a sort of midpoint; half of these states had average annual changes in achievement above the median and half had average annual changes below.[9] The median provides a rough way to compare the magnitude of changes on state tests and NAEP for the entire group of states with sufficient data, including the minority with declines.

As shown in **table 6**, the median increase in the percentage proficient on state tests was larger than the median increase in the percentage basic on NAEP between 2005 and 2009. The same pattern was apparent in effect sizes.

When we calculated the medians in table 6, we also looked at the largest gain and the largest decline found in any state for a particular grade/subject on the state test and on NAEP. As it turned out, the maximum gain on a state test exceeded the maximum gain on NAEP for all grade/subject combinations—sometimes by a very great margin. Even more interesting, the largest declines on state tests were also greater than the largest declines on NAEP, an observation that runs counter to the score inflation argument.

As a final analysis, we sought to determine whether there was a correlation between the size of the gains on state tests and the size of the gains on NAEP by computing statistics called correlation coefficients. In other words, was there evidence to suggest that the larger the gain a state made on its state test, the larger the gain it made on NAEP? For the period from 2005 to 2009, we found weak correlations in most grade/subject combinations in the size of percentage proficient gains on state tests and percentage basic gains on NAEP. Only in grade 8 math was there a moderate degree of correlation in the size of gains. For the period from 2007 through 2009, correlations in percentages proficient/basic were weak to moderate. Correlations in effect sizes for both the longer and shorter time spans were weak to non-existent.

[9] Again, it is important to remember that the median percentage proficient for state tests is based on many different state tests of varying difficulty, whereas the median percentage basic on NAEP is based on the same assessment for all states.

Table 6. Median average yearly gains on state tests and NAEP, 2005 through 2009

Changes in the median average yearly gain in percentages proficient/basic								
	Grade 4 reading		Grade 8 reading		Grade 4 math		Grade 8 math	
Median percentage point gain*	State 0.8	NAEP 0.5	State 1.8	NAEP 0.8	State 1.3	NAEP 0.5	State 1.8	NAEP 1.0
Changes in the median average yearly gain in mean scores								
	Grade 4 reading		Grade 8 reading		Grade 4 math		Grade 8 math	
Median gain in standard deviations*	State 0.02	NAEP 0.01	State 0.06	NAEP 0.01	State 0.04	NAEP 0.02	State 0.05	NAEP 0.04

Table reads: Between 2005 and 2009, the median average annual gain in the percentage proficient on state tests of grade 4 reading was 0.8 percentage point, larger than the median average annual gain in the percentage basic on NAEP of 0.5 percentage point.

*The medians for the average yearly gain in percentages proficient on state tests and percentages basic on NAEP are expressed in terms of percentage point gains. The medians for the average yearly gain in effect sizes are expressed in terms of standard deviations.

On balance, we found little relationship between the size of a gain or decline on state tests and the size of a gain or decline on NAEP.

Conclusion

This study found that NAEP trends from 2005 through 2009 tended to move in the same direction as trends on state tests. To some extent, achievement gains seem to be generalizing to measures other than state tests. An optimistic interpretation is that students have learned more in reading and math since 2005.

Although trends on state tests and NAEP often moved in the same upward direction, gains on state tests tended to be larger than gains on NAEP. The states with the largest gains on their state tests were not the same as the states with the largest gains on NAEP.

Several possible factors may explain the larger gains on state tests:

- **Instruction is more closely aligned to state content standards than to NAEP frameworks.** State tests used for NCLB must be aligned to the state's academic content standards, which in turn drive curriculum and instruction. Ideally, the content of tests should correspond with what is taught in most classrooms, although some states have done a better job of alignment than others. NAEP, however, is not aligned intentionally to any state's standards and therefore may be less instructionally sensitive—in other words, it may not reflect what students are actually being taught. State tests may be more instructionally sensitive than NAEP and therefore reflect larger gains.

- **Score inflation on state tests.** A less optimistic explanation, espoused by many researchers, is that scores on state tests have become inflated as a result of inappropriate teaching to state tests (e.g., Koretz, 2005). In an effort to raise test scores in the easiest way possible, teachers

may engage in narrow test preparation targeted at the specific format and content of state tests. As a result, state tests scores may increase without real, meaningful gains in students' knowledge of the broader domains of reading and math that the test is designed to measure.

- **Motivation.** The state tests used for NCLB have high stakes for educators. Federal and state sanctions for districts and schools are determined largely by the results of these tests. State test scores are reported to parents, published in the media, and accessible online. To avoid the sanctions and negative publicity that low test scores can bring, teachers and administrators often go to great lengths to encourage students to take these tests seriously. NAEP, by contrast, has low stakes for educators and students because it is not connected to any direct rewards or sanctions other than the publicizing of results for the nation and the states. Neither students nor their parents receive any individual NAEP results. Because of the low stakes, students may not be motivated to perform their best on NAEP. While it is not clear how this difference in motivation would affect trends, it is possible that conditions could have changed in ways that affected motivation on a state test, NAEP, or both.

- **Subtle changes in test difficulty.** Our achievement studies for the past three years have excluded states from our analyses if they have officially changed their test or cut scores in ways that affect the year-to-year comparability of test data. Nevertheless, state officials or testing contractors can make informal or subtle decisions about testing programs that effectively make tests easier or more difficult over time, such as changing procedures for choosing and scoring test items, changing how weights are assigned to test items, or not precisely equating test forms from year to year. These unpublicized changes could lead to increases (or decreases) in state test scores.

It is likely that some combination of these factors explains the differences in the size of gains between the two assessments. Different factors may be present in various states or various grade levels. Other factors, such as simple familiarity with the content and format of state tests, may also have a positive effect on scores. It is very difficult to sort out the extent to which differences between state tests and NAEP are attributable to each of these factors.

Interestingly, the largest declines on state tests were greater than the largest declines on NAEP, which is inconsistent with other evidence suggesting that state test scores are inflated. In addition, NAEP gains were greater than state test score gains in a limited number of instances. This may indicate that score inflation is less of a factor in some states than others.

Because it is difficult to sort out the extent to which the aforementioned factors explain the differences between state tests and NAEP, and because the two different types of tests assess different skills and serve different purposes, it is difficult to say which test is the "better" source of information about student achievement. Rather than treating NAEP as "more credible" and state test results as somewhat fictional, we prefer to view them in tandem, as a complement to each other, precisely because of the differences in the two types of assessments. When drawing conclusions about trends in student achievement, it is best to consider both sources of data. To the extent that state and NAEP trends converge within a state, conclusions about changes in student achievement will be more justifiable. To the extent that trends on the two assessments diverge, educators and policymakers will need to be more cautious about drawing conclusions and should explore in more depth why the two measures show conflicting trends.

At the same time, policymakers and the public must also recognize that some state tests may lend themselves better to score inflation than others and that test results in some states may be more trustworthy than others. It is certainly fair to question results in some states that show miraculous increases in state test scores.

References

Bandeira de Mello, V., Blankenship, C., & McLaughlin, D. H. (2009). *Mapping state proficiency standards onto NAEP scales: 2005-2007* (NCES 2010-456). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Center on Education Policy. (2007). *Answering the question that matters most: Has student achievement increased since No Child Left Behind?* Washington, DC: Author.

Center on Education Policy. (2008). *Has student achievement increased since 2002? State test score trends through 2006-07.* Washington, DC: Author.

Center on Education Policy (2009). *Is the emphasis on "proficiency" shortchanging higher- and lower-achieving students?* Washington, DC: Author.

Fuller, B., Gesicki, K., Kang, E., & Wright, J. (2006). *Is the No Child Left Behind Act working? The reliability of how states track achievement.* Retrieved on June 28, 2010, from www.aera.net/uploadedFiles/Publications/Journals/Educational_Researcher/3605/07EDR07_268-278.pdf

Jacob, B. A. (2007). *Test-based accountability and student achievement: An investigation of differential performance on NAEP and state assessments.* Cambridge, MA: National Bureau of Economic Research.

Kane, T. J., & Staiger, D. O. (2002). *V*olatility in school test scores: Implications for test-based accountability systems. In D. Ravitch (Ed.), *Brookings papers on education policy, 2002*, 235–260. Washington, DC: Brookings Institution.

Koretz, D. (2005). *Alignment, high stakes, and the inflation of test scores.* International Project for the Study of Educational Accountability Systems. Retrieved on June 28, 2010, from http://ipea.hmdc.harvard.edu/files/ipea/Koretz_ysse_027.pdf

Linn, R. L., & Haug, C. (2002). *Stability of school building accountability scores and gains* (CSE Tech. Rep. No. 561). Los Angeles, CA: Center for the Study of Evaluation.

Medina, J. (2010, July 19). State's exams become easier to pass, education officials say. *New York Times.* Retrieved on July 29, 2010, from www.nytimes.com/2010/07/20/nyregion/20tests.html?_r=1

National Assessment Governing Board. (n.d.). *Some frequently asked questions about the National Assessment of Educational Progress (NAEP) or The Nation's Report Card.* Retrieved July 2, 2010, from www.nagb.org/faqs.htm.

New York State Department of Education. (2010, July 28). Grade 3-8 math and English test results released: Cut scores set to new college-ready proficiency standards [Press release]. Retrieved on August 3, 2010, from www.oms.nysed.gov/press/Grade3-8_Results07282010.html

New York Daily News. (2010, July 28). Shock to the system: New test scores show how far N.Y. really is from the top [Editorial]. Retrieved on August 3, 2010, from www.nydailynews.com/opinions/2010/07/28/2010-07-28_shock_to_the_system.html

Ravitch, D. (2009, August 14). Toughen the test. *New York Post.* Retrieved on July 10, 2010, from www.nypost.com/p/news/opinion/opedcolumnists/item_EF91FoQ4g70eT5y9CCXV1K;jsessionid=BD70062077631DCF7856B01A52AB0473

Stern, S. (2010). Can New York clean up its testing mess? *City Journal, (20)2*. Retrieved on July 10, 2010, from www.city-journal.org/2010/20_2_ny-education-testing.html

Appendix I. Tables with Specific State Names for Data Included in This Report

Table 1-A. Number (and percentage) of states with gains and declines on state tests and NAEP from 2005 to 2009

Subject, grade, trend	State proficient trend		NAEP basic trend	
Grade 4 reading				
# of states with sufficient data	21	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, NM, OH, TN, TX, UT, WA, WI	21	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, NM, OH, TN, TX, UT, WA, WI
# of states with gains	16 *(76%)*	AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, OH, TX, UT	16 *(76%)*	AK, AL, AZ, CA, CO, FL, IA, MA, MD, MT, ND, NE, NM, OH, TN, TX
# of states with declines	3 *(14%)*	TN, WA, WI	3 *(14%)*	LA, UT, WA
# of states with no change	2 *(10%)*	AK, NM	2 *(10%)*	AR, WI
Grade 8 reading				
# of states with sufficient data	21	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MD, MT, ND, NE, NM, NV, OH, PA, TN, TX, UT, WI	21	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MD, MT, ND, NE, NM, NV, OH, PA, TN, TX, UT, WI
# of states with gains	20 *(95%)*	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MD, MT, ND, NE, NM, NV, PA, TN, TX, UT, WI	17 *(81%)*	AK, AL, AZ, CA, CO, FL, MD, MT, ND, NM, NV, OH, PA, TN, TX, UT, WI
# of states with declines	1 *(5%)*	OH	1 *(5%)*	IA
# of states with no change	0 *(0%)*		3 *(14%)*	AR, LA, NE

(continued)

Table 1-A. Number (and percentage) of states with gains and declines on state tests and NAEP from 2005 to 2009 (Continued)

Subject, grade, trend	State proficient trend		NAEP basic trend	
Grade 4 math				
# of states with sufficient data	19	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, NM, TN, TX, WA, WI	19	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, NM, TN, TX, WA, WI
# of states with gains	18 *(95%)*	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, NM, TN, TX, WI	15 *(79%)*	AK, AL, AR, AZ, CA, CO, FL, IA, MA, MD, MT, ND, NE, NM, WI
# of states with declines	1 *(5%)*	WA	2 *(11%)*	LA, TX
# of states with no change	0 *(0%)*		2 *(11%)*	TN, WA
Grade 8 math				
# of states with sufficient data	21	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, NM, NV, OH, PA, TN, TX, WI	21	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, NM, NV, OH, PA, TN, TX, WI
# of states with gains	20 *(95%)*	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, ND, NE, NM, NV, OH, PA, TN, TX, WI	20 *(95%)*	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NM, NV, OH, PA, TN, TX, WI
# of states with declines	1 *(5%)*	MT	0 *(0%)*	
# of states with no change	0 *(0%)*		1 *(5%)*	NE

Table reads: Of the 19 states with sufficient state test data in grade 4 math, 18 states (95%) showed gains between 2005 and 2009 in the percentage of students reaching the proficient level on state tests. Fifteen of these 19 states (79%) showed gains during this period in the percentage of students reaching the basic level on NAEP.

Table 2-A. Extent of agreement between state tests and NAEP in percentage proficient/basic trends, 2005 to 2009

Subject, grade, trend	State proficient trend		NAEP basic trend	
READING		**Grade 4**		**Grade 8**
Number of states with sufficient data	21	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, NM, OH, TN, TX, UT, WA, WI	21	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MD, MT, ND, NE, NM, NV, OH, PA, TN, TX, UT, WI
Number of states where trends agree	14		16	
Both up	13	AL, AZ, CA, CO, FL, IA, MA, MD, MT, ND, NE, OH, TX	16	AK, AL, AZ, CA, CO, FL, MD, MT, ND, NM, NV, PA, TN,TX, UT, WI
Both down	1	WA	0	
Trends disagree	3	LA, TN, UT	2	IA, OH
One flat	4	AK, AR, NM, WI	3	AR, LA, NE
Percentage of states in agreement	67%		76%	
MATHEMATICS		**Grade 4**		**Grade 8**
Number of states with sufficient data	19	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, NM, TN, TX, WA, WI	21	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, NM, NV, OH, PA, TN, TX, WI
Number of states where trends agree	15		19	
Both up	15	AK, AL, AR, AZ, CA, CO, FL, IA, MA, MD, MT, ND, NE, NM, WI	19	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, ND, NM, NV, OH, PA, TN, TX, WI
Both down	0		0	
Trends disagree	2	LA, TX	1	MT
One flat	2	TN, WA	1	NE
Percentage of states in agreement	79%		90%	

Table reads: Of the 21 states with sufficient data in grade 4 reading, trends in the percentages of students reaching the proficient level on state tests and the basic level on NAEP moved in the same direction between 2005 and 2009 in 14 states. Thirteen of these states made gains on both the state test and NAEP, while one state showed a decline on both assessments. Altogether, trends on state tests and NAEP moved in the same direction in 67% of the states with sufficient data in grade 4 reading.

Table 3-A. Extent of agreement between state tests and NAEP in mean score trends, 2005 to 2009

READING	Grade 4		Grade 8	
Number of states with state and NAEP data	15	AL, AR, AZ, CA, CO, FL, IA, LA, MT, ND, NM, TN, TX, UT, WA	18	AL, AR, AZ, CA, CO, DE, FL, IA, LA, MT, ND, NM, NV, OR, PA, TN, TX, UT
Number of states where trends agree	13		15	
Both up	11	AL, AZ, CA, CO, FL, IA, MT, ND, NM, TN, TX	14	AL, AR, AZ, CA, FL, LA, MT, NM, NV, OR, PA, TN, TX, UT
Both down	2	UT, WA	1	DE
Trends disagree	2	AR, LA	3	CO, IA, ND,
One flat	0		0	
Percentageof states in agreement	87%		83%	
MATHEMATICS	**Grade 4**		**Grade 8**	
Number of states with state and NAEP data	14	AL, AR, AZ, CA, CO, FL, IA, LA, MT, ND, NM, TN, TX, WA	17	AL, AR, AZ, CA, CO, DE, FL, IA, LA, MT, ND, NM, NV, OR, PA, TN, TX
Number of states where trends agree	11		16	
Both up	11	AL, AR, AZ, CA, CO, FL, IA, MT, ND, NM, TN	16	AL, AR, AZ, CA, CO, DE, FL, IA, LA, ND, NM, NV, OR, PA, TN, TX
Both down	0		0	
Trends disagree	3	LA, TX, WA	1	MT
One flat	0		0	
Percentage of states in agreement	79%		94%	

Table reads: Of the 15 states with sufficient data in grade 4 reading, trends in mean scores from 2005 through 2009 moved in the same direction on state tests and NAEP in 13 states. Eleven of these states made gains on both the state test and NAEP, while two states showed a decline on both assessments. Altogether, trends on state tests and NAEP moved in the same direction in 87% of the states with sufficient data in grade 4 reading.

Table 4-A. Extent of agreement on state tests and NAEP, 2007 to 2009

READING	Grade 4		Grade 8	
Percentage proficient/basic				
Number of states with state and NAEP data	43	AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, ID, IL, KS, KY, LA, MA, MD, ME, MI, MN, MO, MT, ND, NE, NH, NM, NV, NY, OH, OR, PA, RI, TN, TX, UT, VA, VT, WA, WI, WY	43	AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, ID, IL, KS, KY, LA, MA, MD, ME, MI, MN, MO, MT, ND, NE, NH, NM, NV, NY, OH, OR, PA, RI, TN, TX, UT, VA, VT, WA, WI, WY
Number in which state and NAEP trends agree	21	AK, CA, CO, CT, DC, FL, KY, MA, MD, MI, MO, NH, NM, NY, OR, RI, TN, VT, WA, WI, WY	23	AL, AZ, CA, CT, DC, FL, GA, HI, IL, KY, MD, MN, MO, ND, NE, NM, NV, PA, RI, TN, UT, WA, WI
Percentage in which state and NAEP trends agree	49%		53%	
Mean scores				
Number of states with state and NAEP data	38	AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, ID, IL, KS, KY, LA, ME, MI, MN, MO, MT, ND, NH, NM, NV, NY, OR, PA, RI, TN, TX, UT, VT, WA, WI, WY	38	AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, ID, IL, KS, KY, LA, ME, MI, MN, MO, MT, ND, NH, NM, NV, NY, OR, PA, RI, TN, TX, UT, VT, WA, WI, WY
Number in which state and NAEP trends agree	22	AK, AZ, CA, CO, CT, DC, FL, LA, MI, MO, ND, NH, NM, NV, NY, OR, RI, TN, UT, VT, WI, WY	25	AL, AR, AZ, CA, CT, DC, DE, FL, GA, HI, IL, LA, MI, MN, MO, ND, NH, NM, NV, NY PA, RI, TN, WI, UT
Percentage in which state and NAEP trends agree	58%		66%	

(continued)

Table 4-A. Extent of agreement on state tests and NAEP, 2007 to 2009 (continued)

MATHEMATICS	Grade 4		Grade 8	
Percentage proficient/basic				
Number of states with state and NAEP data	43	AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, HI, IA, ID, IL, KS, KY, LA, MA, MD, ME, MI, MN, MO, MT, NC, ND, NE, NH, NM, NV, NY, OH, OR, PA, RI, SD, TN, TX, VA, VT, WA, WI, WY	43	AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, HI, IA, ID, IL, KS, KY, LA, MA, MD, ME, MI, MN, MO, MT, NC, ND, NE, NH, NM, NV, NY, OH, OR, PA, RI, SD, TN, TX, VA, VT, WA, WI, WY
Number in which state and NAEP trends agree	15	AK, CA, CO, CT, DC, KY, MD, ME, MN, NC, NE, NH, OR, RI, WY	29	AL, AR, AZ, CO, CT, DC, DE, FL, HI, ID, IL, KY, MD, MI, MN, MO, NC, NE, NH, NM, NV, NY, OR, PA, RI, SD, TN, WA, WI
Percentage in which state and NAEP trends agree	35%		67%	
Mean scores				
Number of states with state and NAEP data	38	AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, HI, IA, ID, IL, KS, KY, LA, ME, MI, MN, MO, MT, NC, ND, NH, NM, NV, NY, OR, PA, RI, SD, TN, TX, VT, WA, WI, WY	38	AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, HI, IA, ID, IL, KS, KY, LA, ME, MI, MN, MO, MT, NC, ND, NH, NM, NV, NY, OR, PA, RI, SD, TN, TX, VT, WA, WI, WY
Number in which state and NAEP trends agree	21	AK, CA, CO, CT, DC, FL, ID, IL, KS, KY, ME, MN, MO, NH, NM, PA, RI, SD, TN, TX, VT	31	AL, AR, AZ, CA, CO, CT, DC, DE, FL, HI, ID, KY, MI, MN, MO, MT, NC, ND, NH, NM, NV, NY, OR, PA, RI, SD, TN, TX, VT, WA, WI
Percentage in which state and NAEP trends agree	55%		82%	

Table reads: Of the 43 states with sufficient data in grade 4 math for 2007 through 2009, trends in the percentage proficient on state tests and in the percentage basic on NAEP moved in the same direction during this period in 15 states, or 35% of these states.

Table 5-A. Number (and percentages) of states in which gains from 2005 through 2009 were larger on state tests or NAEP

READING	Grade 4		Grade 8	
Proficient/basic trend				
Number of states with gains on one or both assessments	19*	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, NM, OH, TN, TX, UT	21*	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MD, MT, ND, NE, NM, NV, OH, PA, TN, TX, UT, WI
State gain › NAEP gain	12 *(63%)*	AR, AZ, CA, LA, MA, MD, MT, ND, NE, OH, TX, UT	16 *(76%)*	AL, AR, AZ, CA, IA, LA, MD, MT, ND, NE, NM, NV, PA, TN, TX, UT
NAEP gain › state gain	6 *(32%)*	AK, AL, CO, FL, NM, TN	4 *(19%)*	AK, CO, OH, WI
Mean score (effect size) trend				
Number of states with gains on one or both assessments	13*	AL, AR, AZ, CA, CO, FL, IA, LA, MT, ND, NM, TN, TX	17*	AL, AR, AZ, CA, CO, FL, IA, LA, MT, ND, NM, NV, OR, PA, TN, TX, UT
State gain › NAEP gain	8 *(62%)*	AR, AZ, CA, IA, LA, MT, ND, TX	14 *(82%)*	AR, AZ, CA, FL, IA, LA, MT, ND, NM, NV, OR, PA, TN, TX
NAEP gain › state gain	4 *(31%)*	AL, CO, FL, NM	2 *(12%)*	CO, UT
MATHEMATICS	**Grade 4**		**Grade 8**	
Proficient/basic trend				
Number of states with gains on one or both assessments	18	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, NM, TN, TX, WI	21*	AK, AL, AR, AZ, CA, CO, FL, IA, LA, MA, MD, MT, ND, NE, NM, NV, OH, PA, TN,TX, WI
State gain › NAEP gain	14 *(78%)*	AK, AL, AR, AZ, CA, FL, LA, MA, MD, MT, NE, TN, TX, WI	16 *(76%)*	AL, AR, CA, FL, IA, LA, MA, MD, ND, NE, NM, NV, OH, PA, TX, WI
NAEP gain › state gain	4 *(22%)*	CO, IA, ND, NM	4 *(19%)*	AK, AZ, MT, TN
Mean score (effect size) trend				
Number of states with gains on one or both assessments	14*	AL, AR, AZ, CA, CO, FL, IA, LA, MT, ND, NM, TN, TX, WA	17	AL, AR, AZ, CA, CO, DE, FL, IA, LA, MT, ND, NM, NV, OR, PA, TN, TX
State gain › NAEP gain	9 *(64%)*	AL, AR, AZ, CA, FL, LA, MT, TN, TX	12 *(71%)*	AL, AR, AZ, CA, DE, IA, LA, NM, NV, PA, TN, TX
NAEP gain › state gain	4 *(29%)*	IA, ND, NM, WA	3 *(18%)*	CO, MT, ND

Table reads: Of the 19 states with gains on at least one assessment (the state test and/or NAEP) in the percentages of students scoring proficient/basic in grade 4 reading, the gain was larger on the state test than on NAEP in 12 states and was larger on NAEP than on the state test in 6 states.

*The numbers below do not add up to the total number of states with gains because some states had the same size gains on the state test and NAEP.

Appendix 2. State-by-State Percentages Proficient/Basic on State Tests and NAEP

Note: In some cases, the 2005 and 2009 percentages for a particular state are listed as the same in the table but are not identical due to rounding; this explains why in some cases a slight average annual gain or decline is shown.

Grade 4 reading:
Percentages proficient/basic on state tests and NAEP, 2005–2009

	State proficient			NAEP basic		
State	2005	2009	Average annual gain	2005	2009	Average annual gain
AK	78%	78%	**0.0**	58%	59%	**0.3**
AL	83%	87%	**0.8**	53%	62%	**2.3**
AR	51%	70%	**4.8**	63%	63%	**0.0**
AZ	65%	72%	**1.8**	52%	56%	**1.0**
CA	47%	61%	**3.5**	50%	54%	**1.0**
CO	86%	87%	**0.3**	69%	72%	**0.8**
FL	71%	74%	**0.8**	65%	73%	**2.0**
IA	79%	81%	**0.5**	67%	69%	**0.5**
LA	64%	72%	**2.0**	53%	51%	**-0.5**
MA	50%	53%	**0.8**	78%	80%	**0.5**
MD	81%	87%	**1.4**	65%	70%	**1.3**
MT	75%	81%	**1.5**	71%	73%	**0.5**
ND	76%	80%	**1.1**	72%	76%	**1.0**
NE	88%	95%	**1.7**	68%	70%	**0.5**
NM	52%	52%	**0.0**	51%	52%	**0.3**
OH	77%	82%	**1.4**	69%	71%	**0.5**
TN	91%	90%	**-0.2**	59%	63%	**1.0**
TX	79%	84%	**1.3**	64%	65%	**0.3**
UT	78%	78%	**0.1**	68%	67%	**-0.3**
WA	80%	72%	**-1.9**	70%	68%	**-0.5**
WI	82%	82%	**-0.1**	67%	67%	**0.0**

Grade 8 reading: Percentages proficient/basic on state tests and NAEP, 2005–2009

	State proficient			NAEP basic		
State	2005	2009	Average annual gain	2005	2009	Average annual gain
AK	80%	82%	**0.4**	70%	72%	**0.5**
AL	70%	75%	**1.3**	63%	66%	**0.8**
AR	57%	71%	**3.5**	69%	69%	**0.0**
AZ	64%	69%	**1.3**	65%	68%	**0.8**
CA	39%	48%	**2.3**	60%	64%	**1.0**
CO	86%	88%	**0.5**	75%	78%	**0.8**
FL	44%	54%	**2.5**	66%	76%	**2.5**
IA	72%	74%	**0.7**	79%	77%	**-0.5**
LA	50%	62%	**3.0**	64%	64%	**0.0**
MD	66%	80%	**3.5**	69%	77%	**2.0**
MT	64%	81%	**4.3**	82%	84%	**0.5**
ND	72%	76%	**1.1**	83%	86%	**0.8**
NE	88%	95%	**1.8**	80%	80%	**0.0**
NM	52%	62%	**2.6**	62%	66%	**1.0**
NV	51%	61%	**2.5**	63%	65%	**0.5**
OH	79%	72%	**-1.7**	78%	80%	**0.5**
PA	64%	81%	**4.1**	77%	81%	**1.0**
TN	88%	93%	**1.3**	71%	73%	**0.5**
TX	83%	93%	**2.5**	69%	73%	**1.0**
UT	77%	83%	**1.5**	73%	78%	**1.3**
WI	85%	85%	**0.1**	77%	78%	**0.3**

Grade 4 math:
Percentages proficient/basic on state tests and NAEP, 2005–2009

	State proficient			NAEP basic		
State	2005	2009	Average annual gain	2005	2009	Average annual gain
AK	69%	74%	**1.3**	77%	78%	**0.3**
AL	74%	79%	**1.3**	66%	70%	**1.0**
AR	50%	78%	**7.0**	78%	80%	**0.5**
AZ	71%	74%	**0.8**	70%	71%	**0.3**
CA	50%	66%	**4.0**	71%	72%	**0.3**
CO	90%	92%	**0.5**	81%	84%	**0.8**
FL	64%	75%	**2.8**	82%	86%	**1.0**
IA	81%	81%	**0.1**	85%	87%	**0.5**
LA	61%	65%	**1.0**	74%	72%	**-0.5**
MA	40%	48%	**2.0**	91%	92%	**0.3**
MD	77%	89%	**3.2**	79%	85%	**1.5**
MT	56%	67%	**2.8**	85%	88%	**0.8**
ND	79%	81%	**0.4**	89%	91%	**0.5**
NE	90%	96%	**1.4**	80%	82%	**0.5**
NM	39%	42%	**0.7**	65%	72%	**1.8**
TN	87%	90%	**0.9**	74%	74%	**0.0**
TX	81%	86%	**1.3**	87%	85%	**-0.5**
WA	61%	52%	**-2.2**	84%	84%	**0.0**
WI	73%	81%	**2.1**	84%	85%	**0.3**

Grade 8 math:
Percentages proficient/basic on state tests and NAEP, 2005–2009

	State proficient			NAEP basic		
State	2005	2009	Average annual gain	2005	2009	Average annual gain
AK	62%	67%	**1.1**	69%	75%	**1.5**
AL	63%	74%	**2.7**	53%	58%	**1.3**
AR	33%	61%	**7.0**	64%	67%	**0.8**
AZ	61%	63%	**0.5**	64%	67%	**0.8**
CA	34%	44%	**2.5**	57%	59%	**0.5**
CO	75%	81%	**1.5**	70%	76%	**1.5**
FL	59%	66%	**1.8**	65%	70%	**1.3**
IA	75%	77%	**0.6**	75%	76%	**0.3**
LA	51%	59%	**2.0**	59%	62%	**0.8**
MA	39%	48%	**2.3**	80%	85%	**1.3**
MD	52%	66%	**3.5**	66%	75%	**2.3**
MT	63%	60%	**-0.8**	80%	82%	**0.5**
ND	65%	71%	**1.4**	81%	86%	**1.3**
NE	85%	92%	**1.7**	75%	75%	**0.0**
NM	24%	42%	**4.6**	53%	59%	**1.5**
NV	49%	55%	**1.5**	60%	63%	**0.8**
OH	60%	71%	**2.6**	74%	76%	**0.5**
PA	63%	71%	**2.1**	72%	78%	**1.5**
TN	87%	90%	**0.7**	61%	65%	**1.0**
TX	61%	79%	**4.5**	72%	78%	**1.5**
WI	74%	78%	**1.2**	76%	79%	**0.8**

Center on Education Policy
1001 Connecticut Avenue, NW, Suite 522
Washington, D.C. 20036
tel: 202.822.8065
fax: 202.822.6008
e: cep-dc@cep-dc.org
w: www.cep-dc.org

www.cep-dc.org