NAEP
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Mapping State Proficiency Standards Onto the NAEP Scales:

Variation and Change in State Standards for Reading and Mathematics, 2005–2009

NCES 2011-458

U.S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR EDUCATION STATISTICS

Contents

Under the 2001 reauthorization of the Elementary and Secondary Education Act (ESEA) of 1965, states develop their own assessments and set their own proficiency standards to measure student achievement. Each state controls its own assessment programs, including developing its own standards, resulting in great variation among the states in statewide student assessment practices. This variation creates a challenge in understanding the achievement levels of students across the United States.

Since 2003, the National Center for Education Statistics (NCES) has supported research that compares the proficiency standards of the National Assessment of Educational Progress (NAEP) with those of individual states. State assessments are placed onto a common scale defined by NAEP scores, which allows states' proficiency standards to be compared not only to NAEP, but also to each other.

NCES has released three earlier reports using state data for reading and mathematics at grades 4 and 8 from 2003, 2005, and 2007. This report highlights the findings of the study from 2009, reporting results using state data from the 2008–09 academic year and the 2009 NAEP grades 4 and 8 reading and mathematics assessments. It also examines the consistency of mapping results over time by comparing the last three NAEP administrations: 2005, 2007, and 2009.

Additional information about this and the previous studies is available at http://nces.ed.gov/nationsreportcard/studies/statemapping/.

Executive Summary

State-level National Assessment of Educational Progress (NAEP) results are an important resource for policymakers and other stakeholders responsible for making sense of and acting on state assessment results. Since 2003, the National Center for Education Statistics (NCES) has supported research that focuses on comparing NAEP and state proficiency standards. By showing where states' standards lie on the NAEP scale, the mapping analyses offer several important contributions. First, they allow each state to compare the stringency of its criteria for proficiency with that of other states. Second, mapping analyses inform a state whether the rigor of its standards, as represented by the NAEP scale equivalent of the state's standard, changed over time. (A state's NAEP scale equivalent is the score on the NAEP scale at which the percentage of students in a state's NAEP sample who score at or above that value matches the percentage of students in the state who score proficient or higher on the state assessment.) Significant differences in NAEP scale equivalents might reflect changes in state assessments and standards or changes in policies or practices that occurred between the years. Finally, when key aspects of a state's assessment or standards remain the same, these mapping analyses allow NAEP to substantiate state-reported changes in student achievement.

The following are the research questions and the key findings regarding state proficiency standards, as they are measured on the NAEP scale.

How do states' 2009 standards for proficient performance compare with one another when mapped onto the NAEP scale?

There is wide variation among state proficiency standards.

- In 2009, as in 2003, 2005, and 2007, using NAEP as common metric, standards for proficient performance in reading and mathematics varied across states in terms of the levels of achievement required. For example, for grade 4 reading, the difference in the level required for proficient performance between the five states with the highest standards and the five with the lowest standards was comparable to the difference between *Basic* and *Proficient* performance on NAEP. The results for reading at grade 8 and mathematics in both grades were similar.

Most states' proficiency standards are at or below NAEP's definition of *Basic* performance.

- In grade 4 reading, 35 of the 50 states included in the analysis set standards for proficiency (as measured on the NAEP scale) that were lower than the scale score for *Basic* performance on NAEP and another 15 were in the NAEP *Basic* range. In grade 8 reading, 16 of 50 states set standards that were lower than the cut-point for *Basic* performance on NAEP and another 34 were in the NAEP *Basic* range.

- In grade 4 mathematics, seven of the 50 states included in the analysis set standards for proficiency (as measured on the NAEP scale) that were lower than the *Basic* performance on NAEP, 42 were in the NAEP *Basic* range, and one in the *Proficient* range. In grade 8 mathematics, 12 of 49 states included in the analysis set standards that were lower than the *Basic* performance on NAEP, 36 were in the NAEP *Basic* range, and one in the *Proficient* range.

How do the 2009 NAEP scale equivalents of state standards compare with those estimated for 2007 and 2005?

While NAEP adopted a revised reading framework in 2009, comparability with earlier assessments was maintained. During the same period, however, some states made changes in their assessments—changes substantial enough that the states indicated comparisons between scores of successive administrations were not possible.

Comparisons between the 2009 mapping results and the 2005 and 2007 mapping results in reading and mathematics at grades 4 and 8 were conducted separately for states that made changes in their testing systems and for those that made no such changes.

For those states that made substantive changes in their assessments between 2007 and 2009 most moved toward more rigorous standards as measured by NAEP.

- When examined across grades 4 and 8 for both reading and mathematics, of the 34 cases where states reported changes in their assessments (9 states in reading and 8 states in mathematics), the rigor of the standards increased in 21 cases, 8 showed no change in their standards, and in 5 cases the rigor of their standards (as measured by NAEP scale equivalents) decreased.

For those states that made substantive changes in their assessments between 2005 and 2009, changes in the rigor of states' standards as measured by NAEP were mixed but showed more decreases than increases in the rigor of their standards.

- When examined across grades 4 and 8 for both reading and mathematics, of the 79 cases where states reported changes in their assessments (17 states in grade 4 reading, 20 in grade 8 reading, 19 in grade 4 mathematics, and 23 in grade 8 mathematics), the rigor of the standards increased in 25 cases, 14 showed no change in their standards, and in 40 cases the rigor of their standards (as measured by NAEP scale equivalents) decreased.

Does NAEP corroborate a state's changes in the proportion of students meeting the state's standard for proficiency from 2007 to 2009? From 2005 to 2009?

Changes in the proportion of students meeting states' standards for proficiency between 2007 and 2009 are not corroborated by the proportion of students meeting proficiency, as measured by NAEP, in at least half of the states in the comparison sample.

- In both subjects, changes in achievement between 2007 and 2009 on the state assessments do not agree with changes as measured by NAEP in the same period in at least half of the 40 states with comparable assessments in both years (22 to 26 depending on the subject and grade). In other words, the state assessment and NAEP reports show changes in percentages of students meeting the state's standard that are significantly different from each other. In most cases (17 to 22 depending on the subject and grade), states' results show more positive changes than NAEP results (larger gains or smaller losses).

Results of comparisons between changes in the proportion of students meeting states' standards for proficiency between 2005 and 2009 and the proportion of students meeting proficiency, as measured by NAEP, were mixed.

- The changes from 2005 to 2009 were mixed. For the two subject areas and grade levels, 16 to 18 states have comparable assessments in 2005 and 2009. In reading at grade 4 and in mathematics at grade 8, the changes in the proportion of students meeting the state's proficiency standard are not significantly different from the changes in the proportion meeting the standard as measured by NAEP in more than half of the states (10 of 17 states and 10 of 16 states, respectively). However, these changes are different from each other in more than half of the states in reading at grade 8 (14 of 18 states) and mathematics at grade 4 (10 of 16 states). In most cases, states' results showed more positive changes (12 of 14 and 8 of 10 states, respectively).

This page intentionally left blank

Introduction

Since 2003, NCES has compared each state's standard for proficient performance in reading and mathematics by mapping each state's standard onto the appropriate NAEP scale. The results of those comparisons have been provided in three earlier reports, using state data for reading and mathematics at grades 4 and 8 from 2003, 2005, and 2007. This report provides highlights of applying the methodology for mapping state proficiency standards onto the NAEP scales using state data from the 2008–09 academic year and the 2009 NAEP grade 4 and 8 reading and mathematics assessments.

By showing where states' standards lie on the NAEP scale, the mapping analyses allow each state to compare the stringency of its criteria for proficiency with that of other states. Also, mapping analyses inform a state whether the rigor of its standards, as represented by the NAEP scale equivalent of the state's standard, changed over time. Significant differences in NAEP scale equivalents might reflect actual changes in state assessments and standards or changes in policies or practices that occurred between the assessment years. Finally, when state and NAEP assessments remain the same over two assessment periods, these mapping analyses allow NAEP to substantiate state-reported changes in student achievement.

The analyses summarized in this report address the following questions:

- How do states' 2009 standards for proficient performance compare with one another when mapped onto the NAEP scale?
- How do the 2009 NAEP scale equivalents of state proficiency standards compare with those estimated for 2007 and 2005?
- Does NAEP corroborate a state's changes in the proportion of the students meeting the state's standard for proficiency from 2007 to 2009? From 2005 to 2009?

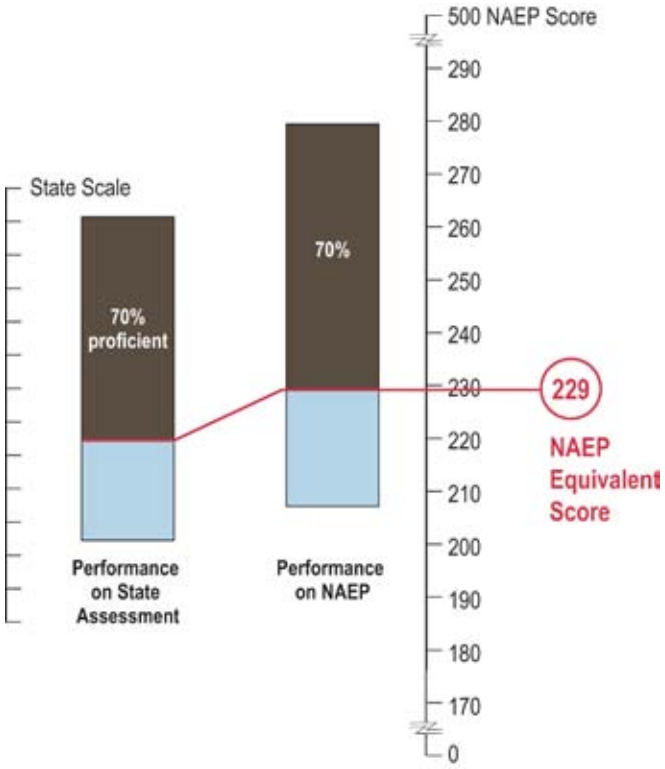
Limitations in the 2003 state assessment data (e.g., many states did not test grades 4 and 8 as NAEP) precluded a 2003 to 2009 comparison analysis.

Mapping of states' standards onto the NAEP scales

The NAEP scale equivalent score corresponding to a state's standard is determined by a direct application of equipercentile mapping. For a given subject and grade, the percentage of students reported in the state assessment to be meeting the standard in each NAEP school is matched to the point in the NAEP achievement scale corresponding to that percentage. In the example depicted in figure 1, if the state reports that 70 percent of the students in fourth grade in a school are meeting their reading achievement standards and 70 percent of the estimated NAEP achievement distribution in that school are at or above 229 on the NAEP scale, then the best estimate from that school's results is that the state's standard is equivalent to 229 on the NAEP scale. These results are aggregated over all of the NAEP schools in a state to provide an estimate of the NAEP scale equivalent of the state's threshold for its standard.

Because states have different standards for proficiency, even if two states report the same percentage of students meeting their own standards, those standards are likely to map onto the NAEP scale at different points (i.e., different states' standards will have different NAEP scale equivalent scores).

Figure 1. Mapping state proficiency standards onto the NAEP scale



The problem with this method is that it could be applied to any set of numbers, whether or not they are meaningfully related. Additional data, beyond the percentage meeting the standard in the state and the distribution of NAEP score—the only data used in the computation—are needed to test the validity of the mapping.

Relative error is a measure of how well the mapping procedure reproduces the percentages reported by the state as meeting the standard in each NAEP-participating school. If the mapping is valid, the procedure should reproduce the individual school percentages fairly accurately, except for some discrepancies emerging from random variation. However, if the state assessment and NAEP are measuring different, uncorrelated characteristics of students, the school-level percentages meeting the state standard as measured by NAEP will bear no relationship to the school-level percentages meeting the state's standards as reported by the state.

The relative error is an indicator of the amount of error that is added to the placement of the standard by the fact that NAEP and the state assessment may not measure the same construct. It is measured as a fraction of the total variation of percentages meeting the standard across schools. When the relative error is greater than .5 (i.e., the mapping error accounts for more than

half of the total variation) then it is considered to be too large to support useful inferences from the placement of the state standard on the NAEP scale without additional evidence.

Additional details on the mapping methodology and relative error are included in the Technical Notes of this report, which can also be found in the 2007 mapping report available at http://nces.ed.gov/nationsreportcard/pdf/studies/2010456.pdf.

Comparisons Over Time

Comparisons between the 2009 mapping results and the 2005 or 2007 mapping results in reading and mathematics at grades 4 and 8 were conducted separately for (a) states that made changes in their testing systems, and (b) those that made no changes. This was done to assess the effects of changes on proficiency standards (for states that made changes), and to find the extent to which NAEP corroborated the changes in achievement measured in the states' assessments between the two periods (for states that did not make changes).

In 2009, a new NAEP reading framework was used in the assessment replacing the framework used through 2007. However, results from special analyses determined that 2009 reading assessment results could be compared with those from earlier assessment years (for more information see http://nces.ed.gov/nationsreportcard/reading/trend_study.asp). In the years from 2005 to 2009, the focus of this report, many states changed their state assessments to ensure that they were complying with the Elementary and Secondary Education Act. Thus, finding differences in their standards is expected. A survey designed to provide contextual information about state assessment programs was conducted. States were asked to indicate, among other things, whether there were significant changes to the state assessment between 2006–07 and 2008–09 affecting the comparability of results.

Data Sources

The analyses in this report are based on NAEP and state assessment results for public schools that participated in the grade 4 and grade 8 NAEP assessments in reading and mathematics, weighted to represent the states. The analyses use data from (a) NAEP data files for the states participating in the 2005, 2007, and 2009 reading and mathematics assessments, (b) state assessment school-level files compiled in the National Longitudinal School-Level State Assessment Score Database (http://www.schooldata.org), and (c) school-level achievement data for the 2006–07 and 2008–09 school years from EDFacts (http://www.ed.gov/EDFacts/), a U. S. Department of Education initiative that centralizes performance data supplied by K-12 state education agencies with other data assets within the Department. The NAEP data used in this report are based on the administration of NAEP assessments to a sample of students from selected public schools in each state, in grades 4 and 8. The files include NAEP achievement data for each selected student. Because state assessment data are only available at the school level, as an initial step in the analysis, NAEP data are aggregated to the school level as well. These school-level data are then aggregated to the state-level taking into account the number of students in the grade at the school. Additional information on sampling and weighting that NAEP uses will be found at http://nces.ed.gov/nationsreportcard/tdw.

The report also relies on a survey of state assessment programs conducted to gain contextual information about the general characteristics of state assessment programs, and to identify changes in states' assessments between the 2004–05 and 2006–07 school years and between the 2006–07 and 2008–09 school years that could affect the interpretation of the mapping study results. The survey was conducted by the NAEP State Coordinators in every state. The survey methodology and summary results for each state are available at http://nces.ed.gov/nationsreportcard/studies/statemapping/.

Cautions in Interpretation

As the earlier mapping reports pointed out (McLaughlin et al. 2008a, 2008b; National Center for Education Statistics 2007; Bandeira de Mello, Blankenship, and McLaughlin 2009), the mapping methodology has several caveats that need to be noted. The methodology does not allow linking scores of individual students on the two tests; the results of this study cannot be used, for example, to map a student's score onto a test score in a second state. This report is not an evaluation of state assessments. State assessments and NAEP are developed for different purposes and have different goals and they may vary in format and administration. Findings of different standards, different trends, and different gaps are presented without suggestion that they be considered as deficiencies either in state assessments or in NAEP. The analyses in this report do not address questions about the content, format, exclusion criteria, or conduct of state assessments, as compared to NAEP. State assessments and their associated proficiency standards are designed to provide pedagogical information about individual students to their parents and teachers, whereas NAEP is designed to provide performance information at an aggregate level. Also, the analyses do not address any change in states' assessments or proficiency standards that may have occurred after 2009.

Mapping the various state proficiency standards on the NAEP scale and comparing the standards with NAEP achievement levels gives context to the discussion, but it does not imply that the NAEP achievement levels are more valid than the state standards or that states should emulate NAEP standards. There is a wide range of policy considerations involved in setting achievement standards, and what is appropriate for NAEP may not be the best fit for a given state. NAEP's achievement levels are used to interpret the meaning of the NAEP scales. NCES has determined (as indicated by NAEP's authorizing legislation) that NAEP achievement levels should continue to be used on a trial basis and should be interpreted with caution. Additional information on NAEP achievement levels are available at http://nces.ed.gov/nationsreportcard/achlevdev.asp.

Steps have been taken to reduce the impact of some of these concerns. For example, the analyses of changes in student achievement over time are made only for state assessments considered comparable after an extensive evaluation of state assessment practices. Regardless of its limitations, this and previous mapping studies provide valuable information in helping understand the myriad of state assessment results, and serves a policy need for reliable information that compares states' standards.

In the report, findings are reported based on a statistical significance level set at .05. When comparisons are made, terms like *decreased* or *increased* indicate statistically significant findings. In all figures, a black triangle next to state names indicates that the relative error is greater than .5.

Organization of This Report

The remainder of this report presents first the analyses that examined the mapping results for 2009 in reading and mathematics at grades 4 and 8. These are followed by the analyses comparing the 2009 NAEP scale equivalents of state standards with those estimated for 2007 and 2005. Finally, we discuss the NAEP and state assessment changes in achievement from 2005 and 2007 to 2009. The last section consists of an appendix that contains relevant technical notes and tables with results complementing those discussed in the body of the report.

State Performance Standards

The analyses in this section address the question of how states' 2009 standards for proficient performance compare with one another when mapped on the NAEP scale. A number of general statements can be made:

- Using NAEP as a common yardstick allows a comparison of different state assessments, which have unique criteria for determining proficiency.
- The range of state standards continues to be wide: 60 to 71 NAEP points, depending on grade and subject. With such a wide range, a student considered proficient in one state may not be considered proficient in another.
- Almost all state standards (50 in grades 4 and 8 reading and in grade 4 mathematics, and 49 in grade 8 mathematics) are mapped at NAEP's *Basic* achievement level or below, which represents partial mastery of knowledge and skills fundamental for proficient work at each grade.
- For grade 4 reading, most state standards (35 of 50) are below the NAEP *Basic* achievement level. For the other three subject-grade combinations, most state standards are within the *Basic* range (33 of 50 in grade 8 reading, 42 of 50 in grade 4 mathematics, and 33 of 49 in grade 8 mathematics).

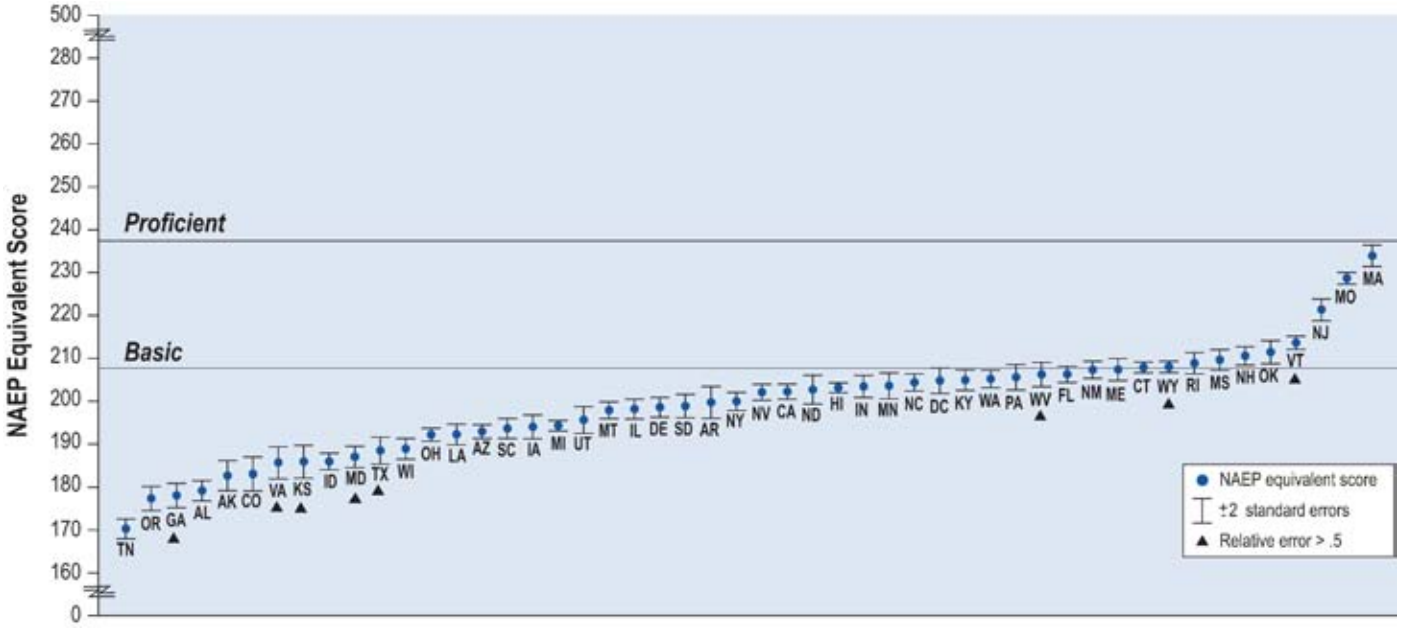
According to the National Assessment Governing Board, students who perform at the *Basic* achievement level show "partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade." Students who perform at the *Proficient* achievement level demonstrate competency over challenging subject matter.

Reading—Grade 4

For grade 4 reading, the NAEP cut point for *Basic* performance is set at 208 and the cut point for *Proficient* at 238. The average across states of the estimated standards for proficient on the NAEP scale was equivalent to a NAEP score of 199, below NAEP's definition of *Basic*. Figure 2 shows each state and its NAEP equivalent score for grade 4 reading. The lines in the figure indicate the cut points for the NAEP *Proficient* and *Basic* performances.

Relative error is a measure of how well the mapping procedure reproduces the percentages reported by the state as meeting the standard in each NAEP-participating school. In figure 2, the black triangle under the state abbreviation indicates that the relative error of the NAEP equivalent of that state's standards is too large to support useful inferences without additional evidence. A more detailed discussion about the relative error is available in the Technical Notes.

Figure 2. NAEP scale equivalents of state grade 4 reading standards for proficient performance, by state: 2009

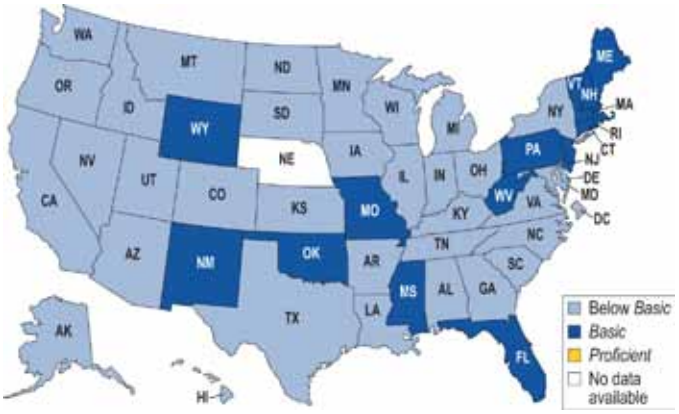


▲ Inferences based on estimates with relative error greater than .5 may require additional evidence.

Although some states in figure 2 have point estimates of their NAEP scale equivalents that are below the cut point for *Basic* performance (208), because of the error associated with the estimate their NAEP scale equivalent may not be significantly different from the cut point. Pennsylvania, Florida, and New Mexico are examples. Therefore, accounting for the margin of error, 35 of 50 states set grade 4 standards for reading proficiency that were lower than the *Basic* performance on NAEP. The remaining states were within the *Basic* range. The difference between the lowest and highest states, Tennessee and Massachusetts, was 64 points.

Figure 3 shows the 35 states whose proficiency standards were below *Basic* and the 15 whose standards were within the *Basic* range.

Figure 3. States' proficiency standards for grade 4 reading classified into NAEP achievement levels: 2009



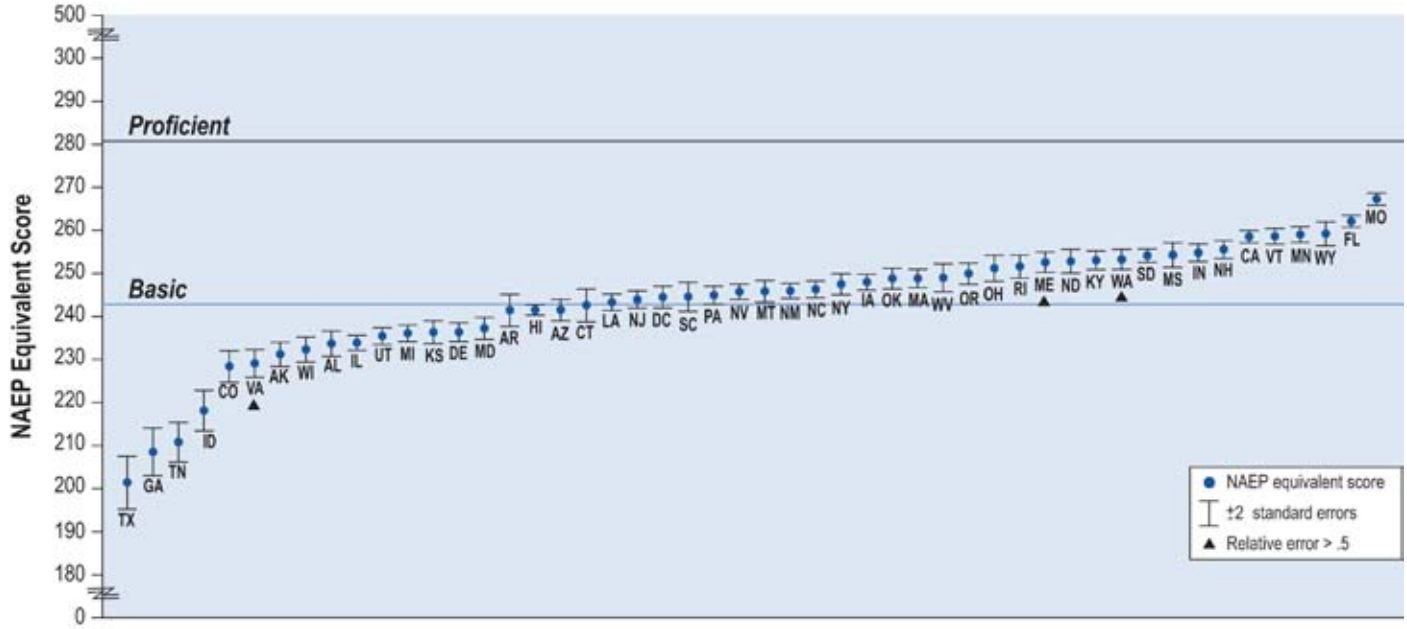
NOTE: In Nebraska, each district develops local assessments to report on standards. Therefore, the state was not included in the analyses.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Reading—Grade 8

For grade 8 reading, NAEP sets the cut point for *Basic* performance at 243 and for *Proficient* at 281. Figure 4 shows each state and its NAEP equivalent score for grade 8 reading. The average state's NAEP equivalent standard for proficiency was 243, at NAEP's definition of *Basic*. Accounting for the margin of error, 16 of 50 states set grade 8 standards for proficiency (as measured on the NAEP scale) that were lower than the *Basic* cut point on NAEP. Not one state had a standard in the *Proficient* range. There was also wide variation between state standards: the range between the lowest state, Texas, and the highest, Missouri, was 66 points.

Figure 4. NAEP scale equivalents of state grade 8 reading standards for proficient performance, by state: 2009



▲ Inferences based on estimates with relative error greater than .5 may require additional evidence.

Figure 5 shows the 16 states whose proficiency standards were below *Basic* and the 34 whose standards were within the *Basic* range.

Figure 5. States' proficiency standards for grade 8 reading classified into NAEP achievement levels: 2009

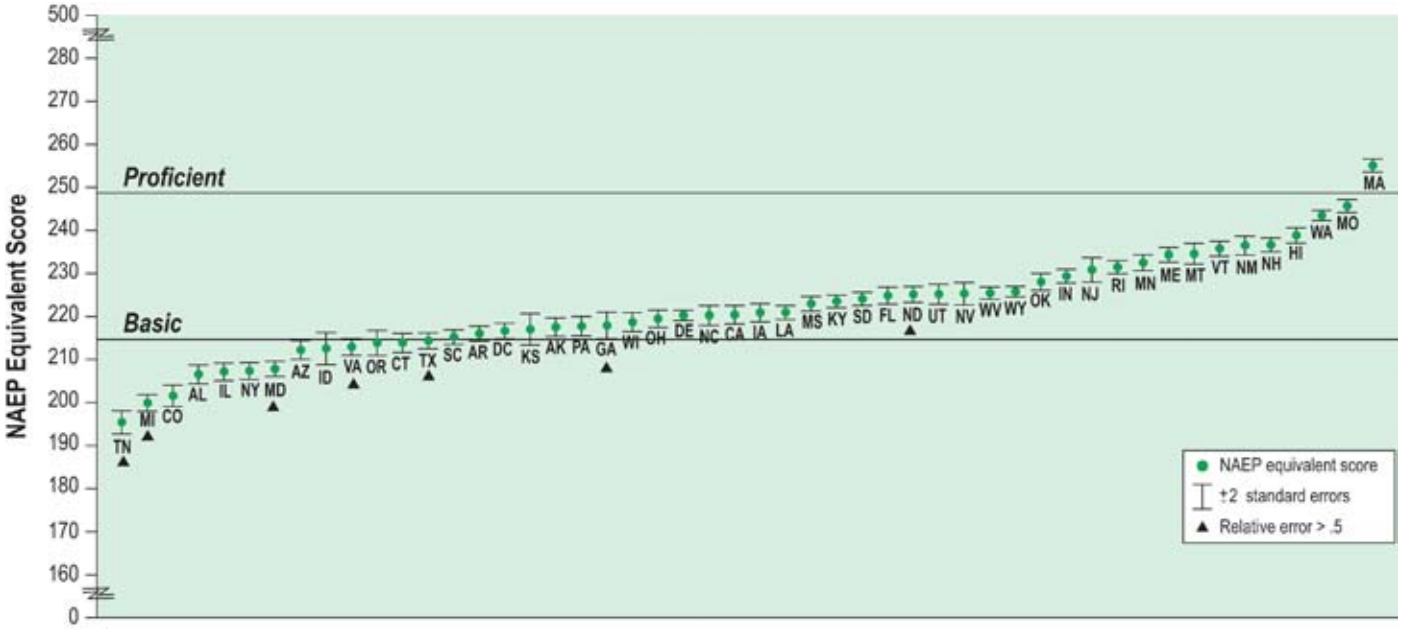
NOTE: In Nebraska, each district develops local assessments to report on standards. Therefore, the state was not included in the analyses.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Mathematics—Grade 4

For grade 4 mathematics, the NAEP cut point for *Basic* performance is 214 and the cut point for *Proficient* is 249. The average NAEP scale equivalent for proficient across the states was 222, within the NAEP *Basic* range. Figure 6 shows the NAEP equivalent mathematics scores for each state for grade 4, including markers for the NAEP *Basic* and *Proficient* standards. Seven of 50 states set grade 4 standards for proficient below the NAEP *Basic* level, and one state set its standard higher than NAEP's *Proficient*. The remainder fell in NAEP's *Basic* range. The variation between the lowest state, Tennessee, and the highest, Massachusetts, was 60 points.

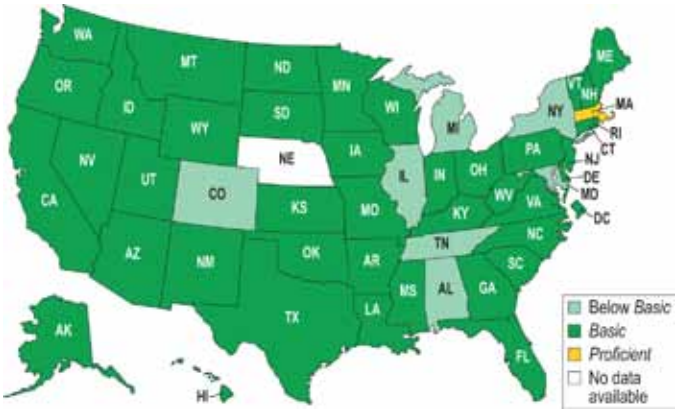
Figure 6. NAEP scale equivalents of state grade 4 mathematics standards for proficient performance, by state: 2009



▲ Inferences based on estimates with relative error greater than .5 may require additional evidence.

Figure 7 shows the seven states whose proficiency standards were below *Basic*, the 42 whose standards were within the *Basic* range, and the one state above the *Proficient* cut point.

Figure 7. States' proficiency standards for grade 4 mathematics classified into NAEP achievement levels: 2009



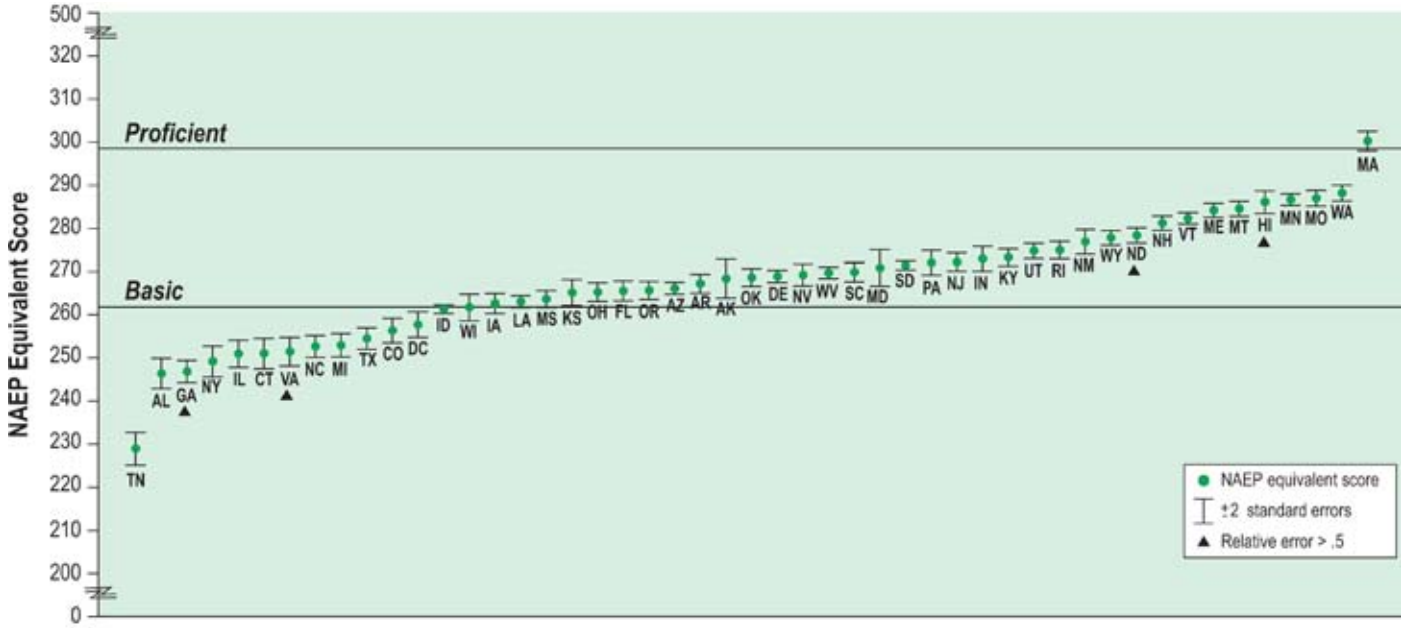
NOTE: In Nebraska, each district develops local assessments to report on standards. Therefore, the state was not included in the analyses.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Mathematics—Grade 8

For grade 8 mathematics, the NAEP cut point for *Basic* is 262 and the cut point for *Proficient* is 299. The average NAEP scale equivalent for state standards was 268, between the NAEP standards of *Basic* and *Proficient*. Figure 8 shows that 12 out of 49 states set grade 8 standards for proficient in mathematics that were lower than *Basic* performance on NAEP, and one state set standards above NAEP's standard of *Proficient*. The difference between the lowest and highest states, Tennessee and Massachusetts, was 71 points.

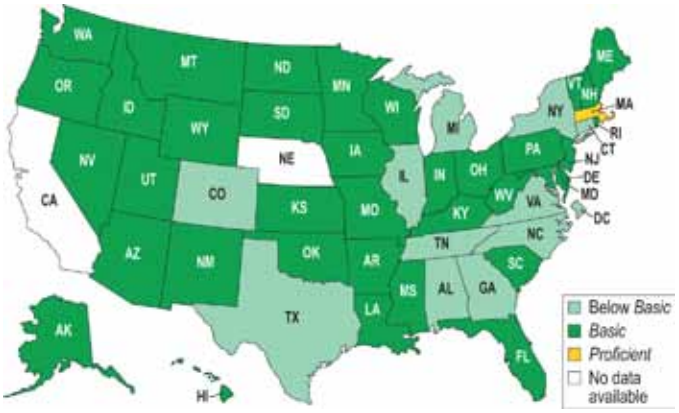
Figure 8. NAEP scale equivalents of state grade 8 mathematics standards for proficient performance, by state: 2009



▲ Inferences based on estimates with relative error greater than .5 may require additional evidence.

Figure 9 shows the 12 states whose proficiency standards were below *Basic*, the 36 whose standards were within the *Basic* range, and the one state above the *Proficient* cut point.

Figure 9. States' proficiency standards for grade 8 mathematics classified into NAEP achievement levels: 2009



NOTE: In Nebraska, each district develops local assessments to report on standards. Therefore, the state was not included in the analyses. California was not included because the state does not test general mathematics.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

State Standards and NAEP Achievement Levels

Figures 10 and 11 show a summary of the state proficiency standards for both reading and mathematics expressed in terms of NAEP achievement levels. In grade 4 reading, as shown in figure 10, all state proficiency standards (as measured by NAEP) fell in the NAEP *Basic* or below *Basic* range. In grade 4 mathematics, most state standards (42 of 50) were within the *Basic* range. For 28 states, their mathematics standards were in the *Basic* range, whereas their reading standards were in the below *Basic* range. For seven states, the grade 4 reading and mathematics proficiency standards fell below the *Basic* range.

Figure 11 shows that the majority of states' grade 8 standards fell within the NAEP *Basic* range for both reading and mathematics (most grade 4 standards fell below *Basic*). Still, eight states had proficiency standards that were below *Basic* for both reading and mathematics, five of which were also below *Basic* for both reading and mathematics in grade 4.

Figure 10. States' proficiency standards for grade 4 reading and mathematics classified into NAEP achievement levels: 2009

		Reading			
		Below *Basic*	*Basic*	*Proficient*	Total
Mathematics	*Proficient*	— 0	MA 1	— 0	1
	Basic	AK, AR, AZ, CA, DC, DE, GA, HI, IA, ID, IN, KS, KY, LA, MN, MT, NC, ND, NV, OH, OR, SC, SD, TX, UT, VA, WA, WI 28	CT, FL, ME, MO, MS, NH, NJ, NM, OK, PA, RI, VT, WV, WY 14	— 0	42
	Below *Basic*	AL, CO, IL, MD, MI, NY, TN 7	— 0	— 0	7
	Total	35	15	0	50

Figure 11. States' proficiency standards for grade 8 reading and mathematics classified into NAEP achievement levels: 2009

		Reading			
		Below *Basic*	*Basic*	*Proficient*	Total
Mathematics	*Proficient*	— 0	MA 1	— 0	1
	Basic	AK, DE, HI, ID, KS, MD, UT, WI 8	AR, AZ, FL, IA, IN, KY, LA, ME, MN, MO, MS, MT, ND, NH, NJ, NM, NV, OH, OK, OR, PA, RI, SC, SD, VT, WA, WV, WY 28	— 0	36
	Below *Basic*	AL, CO, GA, IL, MI, TN, TX, VA 8	CT, DC, NC, NY 4	— 0	12
	Total	16	33	0	49

— No states in the category.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 Reading and Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Similarity of State Assessments and NAEP

A measure of the appropriateness of the mapping is the correlation coefficient showing the relationship between the percentages reported for schools by the state and those estimated from the NAEP scale equivalents: the two assessments must agree on which schools are high achieving and which are not. For each subject and grade, table 1 displays the range of correlations between the school-level percentages meeting the state proficient standard and the percentage of the NAEP sample at or above the NAEP equivalent score in those schools.

Many of the states included in the analyses had state assessment results that were highly correlated with NAEP. Across both subjects and grades, the majority of cases had a correlation of .7 or higher between NAEP and state assessment school-level percentages meeting the proficient standards for grades 4 and 8 reading and mathematics. For those states, both assessments identified similar patterns of achievement across schools. In reading, 52 percent of states at grade 4 and 44 percent of states at grade 8 had correlations of .7 or higher. Correlations were higher in mathematics than in reading: 58 percent of states at grade 4 and 69 percent of states at grade 8 had correlations of .7 or above.

The lower correlations in some states need to be considered when interpreting the comparisons of NAEP and state assessment results. These low correlations could be the result of, for example, small enrollments in these states' schools that affect the reliability of results, or tests that measure different knowledge areas.

Table 1. Frequency of correlations between NAEP and state assessment school-level percentages meeting the proficient standards for reading and mathematics, grades 4 and 8: 2009

	Reading		Mathematics	
Correlation range	Grade 4	Grade 8	Grade 4	Grade 8
Total states[1]	50	50	50	49
.3 ≤ r < .4	1	1	0	1
.4 ≤ r < .5	2	3	5	0
.5 ≤ r < .6	8	9	4	1
.6 ≤ r < .7	13	15	12	13
.7 ≤ r < .8	22	12	27	20
r ≥ .8	4	10	2	14

[1]Nebraska did not have a statewide assessment and was not included in these analyses. California does not test general mathematics in grade 8.

NOTE: Correlations are available by state at http://nces.ed.gov/nationsreportcard/studies/statemapping/.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 Reading and Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Comparing 2009 With 2007 and 2005 State Performance Standards Using NAEP Equivalent Scores

The analyses in this section address the question of how the 2009 NAEP scale equivalents of state standards compared with those estimated for 2007 and 2005. This section compares the states that indicated they made substantive changes in their testing systems during the two periods. By comparing them we can assess the effects of such changes on the states' proficiency standards. The analyses showed the following:

- From 2007 to 2009, there were significant changes in the rigor of state standards as measured by NAEP and most states with significant changes moved to more rigorous standards.
- From 2005 to 2009, there were also significant changes in the rigor of state standards as measured by NAEP; some state standards increased in rigor while others decreased.

The analyses in this section focus on the consistency of mapping outcomes over time using 2005, 2007 and 2009 assessments. In 2009 states still had wide variation in the stringency of their standards. Table 2 shows the difference between the highest and lowest levels of state proficiency standards as measured by the NAEP reading and mathematics scale by grade, for each year of analysis. The smallest gap is 56 points, at grade 4 mathematics in both 2005 and 2007.

Although the NAEP assessments did not change between 2005 and 2009, some states made changes in their state assessments in the same period that were substantial enough that states indicated that comparisons between scores of successive administrations were not possible. Table 3 shows that for each of the four assessments, at least eight states reported that they changed key aspects of the assessment between 2007 and 2009, either modifying the assessment or changing the standard itself. Between 2005 and 2009, at least 17 states reported that they made changes. Tables in appendix list the states by whether they made changes in their assessments in these two periods.

Comparisons between the 2009 and previous mappings were made separately for states that made changes in their testing systems and for those that made no such changes. This section focuses on the states that made changes to their assessments and on the effects those changes had on their proficiency standards.

The mapping can be used to test whether changes in the assessment or in the standard affected the rigor of the standard. Figures 11, 12, 13, and 14 depict the effects of the changes.

Table 2. Differences between the highest and lowest levels of state proficiency standards as measured on the NAEP reading and mathematics scales, grades 4 and 8, by year: 2005, 2007, and 2009

	Reading		Mathematics	
Year	Grade 4	Grade 8	Grade 4	Grade 8
2009	64	66	60	71
2007	69	70	56	78
2005	74	62	56	81

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005, 2007, and 2009 Reading and Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Table 3. Number of states that did or did not make substantive changes in their assessments that affected comparability of results between 2005 and 2009 and between 2007 and 2009

2005 and 2009		Substantive changes	No substantive changes	Total
Reading	Grade 4	17	17	34
	Grade 8	20	18	38
Mathematics	Grade 4	19	16	35
	Grade 8	23	16	39

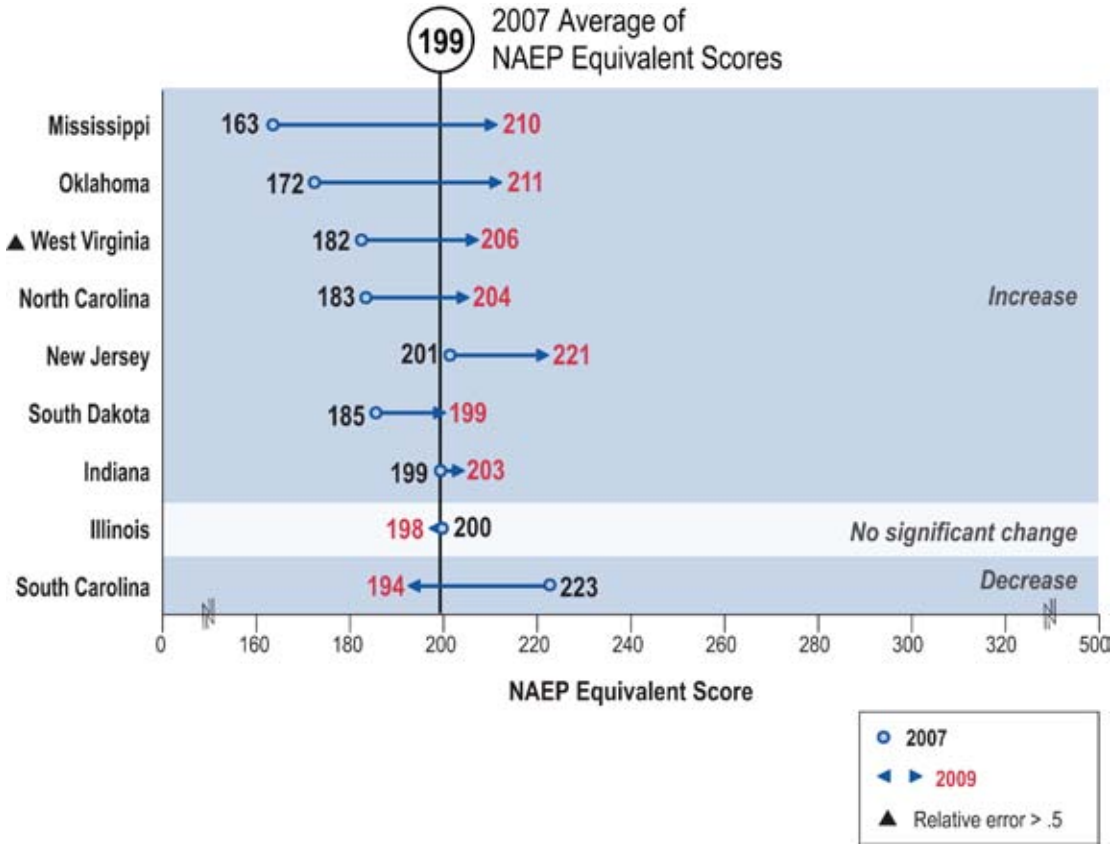
2007 and 2009		Substantive changes	No substantive changes	Total
Reading	Grade 4	9	40	49
	Grade 8	9	40	49
Mathematics	Grade 4	8	41	49
	Grade 8	8	40	48

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 Survey of State Assessment Program Characteristics.

Nine states made substantive changes in their grade 4 reading assessment from 2007 to 2009 (figure 12). Among these states, seven increased the rigor of their reading standards. The 2007 score is shown in black and the 2009 score is shown in red. The arrows point in the direction of the change. For example, Mississippi's reading grade 4 NAEP equivalent score rose from 163 in 2007 to 210 in 2009. The NAEP equivalent score for Illinois did not change significantly whereas for South Carolina the NAEP equivalent score decreased, with the arrowhead pointing to the left.

The average equivalent score for the 49 state proficiency standards for grade 4 reading in 2007 was 199. This average does not reflect a consensus or a goal that all the states should be moving to; it just provides a reference of where these states are in comparison to the average.

Figure 12. Change in the estimated NAEP scale equivalent scores of grade 4 reading proficiency standards for states that made substantive changes in their assessments: 2007 and 2009



▲ Inferences based on estimates with relative error greater than .5 may require additional evidence.

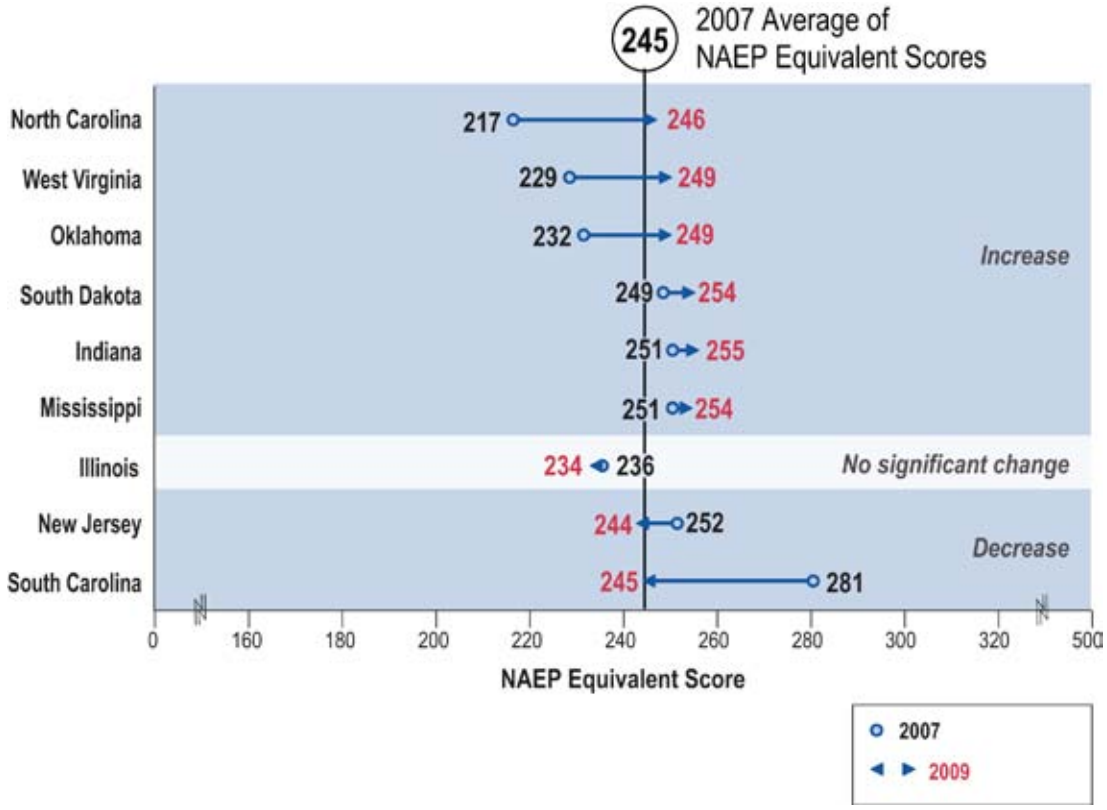
NOTE: The 2007 average of NAEP equivalent scores is based of 49 state standards. State assessment data for the District of Columbia and Nebraska were not available.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 and 2009 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

At grade 8, the same nine states made changes in the reading assessment from 2007 to 2009. With the exception of New Jersey, the states with increased grade 4 proficiency standards also increased the rigor of their grade 8 standards (figure 13). The NAEP equivalent score for Illinois did not change significantly, whereas South Carolina's and New Jersey's NAEP equivalent scores decreased, with both arrows pointing to the left.

Based on 49 states, the average of the state proficiency standards for grade 8 reading in 2007 was 245.

Figure 13. Change in the estimated NAEP scale equivalent scores for grade 8 reading proficiency standards for states that made substantive changes in their assessments: 2007 and 2009

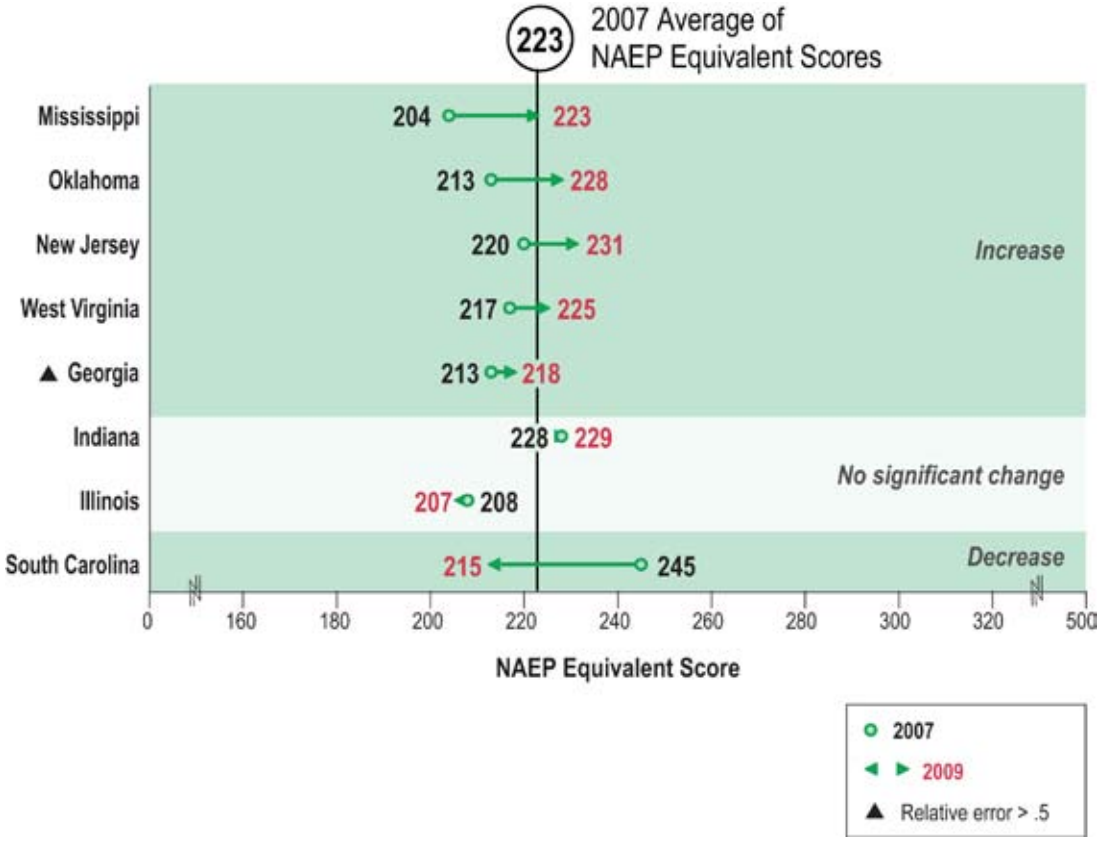


NOTE: The 2007 average of NAEP equivalent scores is based of 49 state standards. State assessment data for the District of Columbia and Nebraska were not available.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 and 2009 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Eight states made changes in their grade 4 mathematics assessments from 2007 to 2009. In five states, the state proficiency standard increased (figure 14). In two states, it did not change significantly. In one state, South Carolina, it decrease significantly. The average of the 49 state proficiency standards for grade 4 mathematics in 2007 was 223.

Figure 14. Change in the estimated NAEP scale equivalent scores for grade 4 mathematics proficiency standards for states that made substantive changes in their assessments: 2007 and 2009



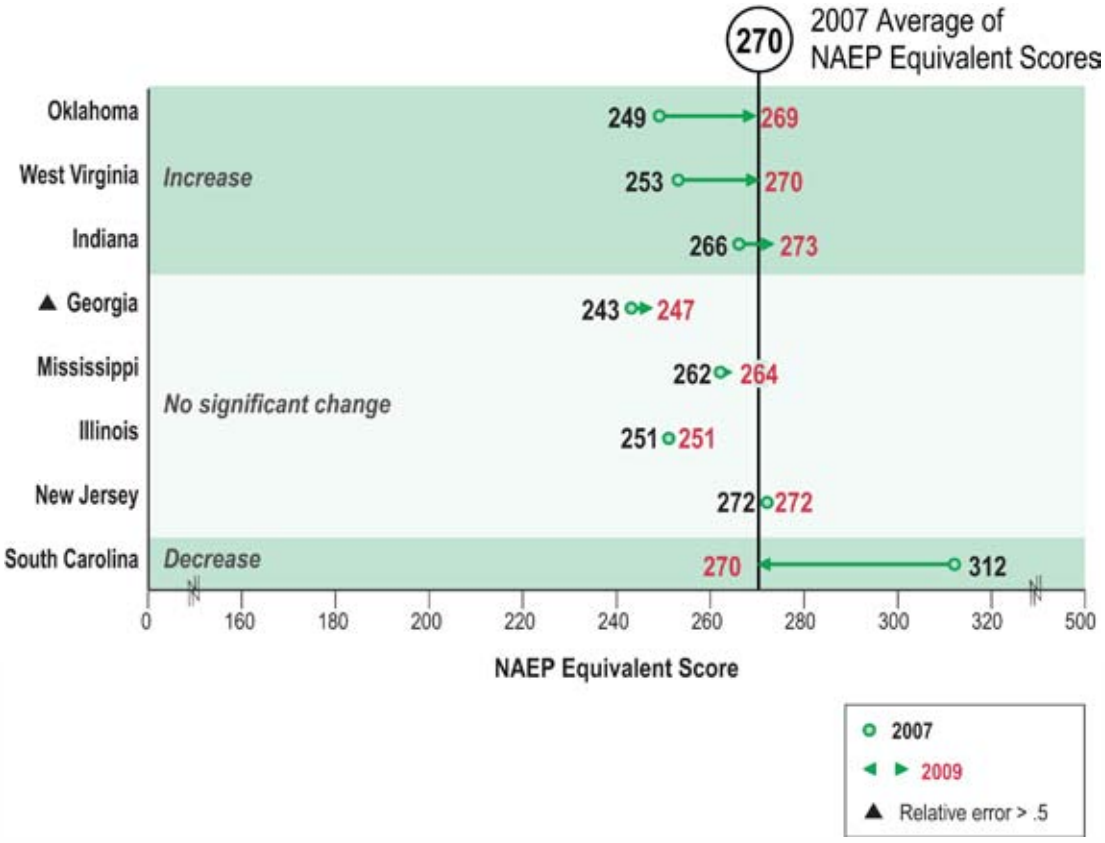
▲ Inferences based on estimates with relative error greater than .5 may require additional evidence.

NOTE: The 2007 average of NAEP equivalent scores is based of 49 state standards. State assessment data for the District of Columbia and Nebraska were not available

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 and 2009 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Eight states made changes in their grade 8 mathematics assessments from 2007 to 2009. The state proficiency standard increased for three of these states. In four states, the state standard did not change significantly. In one state, South Carolina, it decreased significantly. The average of the 48 state proficiency standards for grade 8 mathematics in 2007 was 270.

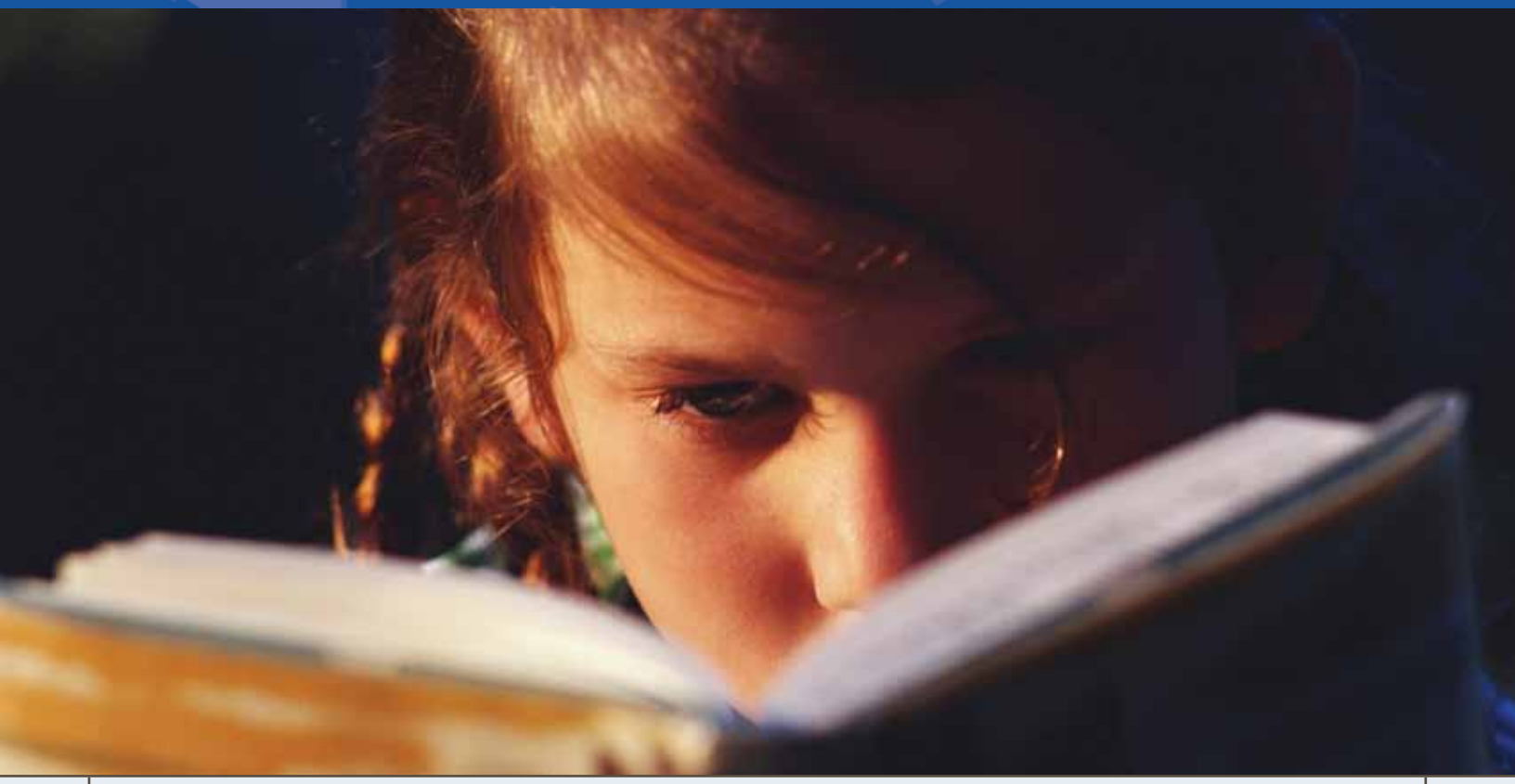
Figure 15. Change in the estimated NAEP scale equivalent scores for grade 8 mathematics proficiency standards for states that made substantive changes in their assessments: 2007 and 2009



▲ Inferences based on estimates with relative error greater than .5 may require additional evidence.

NOTE: The 2007 average of NAEP equivalent scores is based of 48 state standards. State assessment data for California, the District of Columbia, and Nebraska were not available.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 and 2009 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.



Using NAEP to Corroborate State Measures of Achievement Change

In this section, we compare the change over time in the percentages of students meeting a state's standard with the change in the percentages of students meeting the NAEP equivalent of the same state's standard. Comparisons of state assessments over time are possible only when the assessments that were given in 2005, 2007, and 2009 did not meaningfully change. For the two subject areas and grade levels, between 16 and 18 states had comparable assessment data for 2005 and 2009, and 40 and 41 states had comparable assessment data for 2007 and 2009 (table 3).

- Changes in the proportion of students meeting states' standards for proficiency between 2007 and 2009 are not corroborated when compared with the proportion of students meeting proficiency as measured by NAEP. Further, most states show more positive changes (e.g., larger gains or smaller losses) in the proportion meeting the state standards than are shown to meet proficiency when using NAEP.
- Changes in achievement between 2005 and 2009 in state tests are not corroborated by changes in achievement measured by NAEP.

Summary

Looking at the states that made significant changes in an assessment compared with 2007, across grades 4 and 8 for both reading and mathematics, we see that in 21 cases the change resulted in a higher standard, 5 cases showed a decrease, and 8 demonstrated no significant change in the standard.

Table 4 summarizes the results in the previous figures as well as the changes from 2005 to 2009. As can be seen, from 2007 to 2009, increases in standards were more common than decreases, while from 2005 to 2009, the changes were more mixed. The *No significant change* column shows that in many states a change in the assessment did not affect the state's proficiency standard.

Regardless of whether state and NAEP assessments remain the same over two assessment periods, when NAEP scale equivalents are significantly different, further investigations can help establish the factors that may have contributed to such difference. For example, if state assessments remained the same over the comparison period, differences in NAEP scale equivalents could be attributed to changes in instructional practices or curricula placing more emphasis on subject matter covered more on the state test than on NAEP from one assessment year to the next. Also, changes in state exclusion policies might have changed the rates of participation of students with disabilities and/or English language learners in the NAEP or state assessments.

Table 4. Direction of change in the estimated NAEP scale equivalent scores of state proficiency standards for the states that made substantive changes in their assessments, by subject and grade: 2005 to 2009, 2007 and 2009

Reading

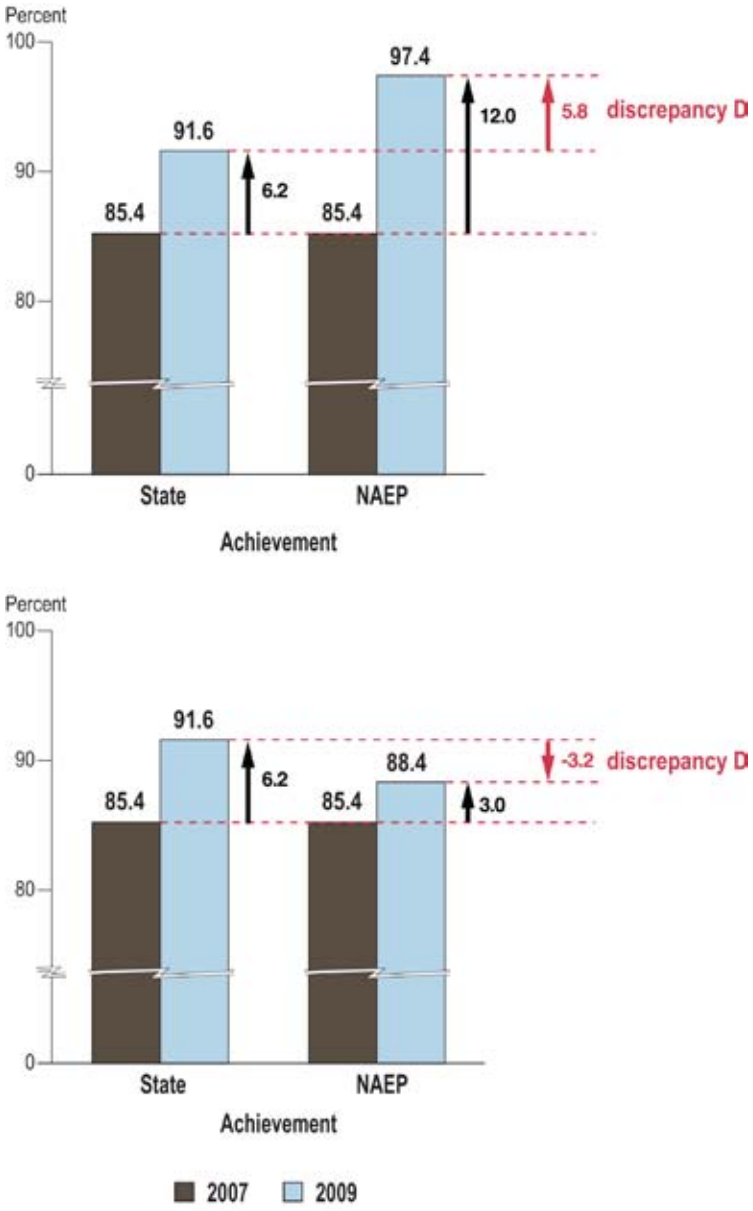
Period		Increase	No significant change	Decrease
2007 to 2009	Grade 4	IN, MS, NC, NJ, OK, SD, WV 7	IL 1	SC 1
	Grade 8	IN, MS, NC, OK, SD, WV 6	IL 1	NJ, SC 2
2005 to 2009	Grade 4	IN, MI, MS, NC, NJ, OK, WV 7	GA, HI, ID, KY, MT 5	CT, ME, NY, SC, WY 5
	Grade 8	IN, MS, NC, OK, WV 5	CT 1	DE, GA, HI, ID, IL, KS, ME, MT, NJ, NY, OR, SC, VA, WY 14

Mathematics

Period		Increase	No significant change	Decrease
2007 to 2009	Grade 4	GA, MS, NJ, OK, WV 5	IN, IL 2	SC 1
	Grade 8	IN, OK, WV 3	GA, IL, MS, NJ 4	SC 1
2005 to 2009	Grade 4	IN, MO, MS, MT, NC, NJ, OK, WV 8	GA, ID, KS, NY 4	CT, HI, ME, MI, OH, SC, WY 7
	Grade 8	IN, MT, NC, OK, WV 5	MA, MS, NJ, VA 4	CT, DE, GA, HI, ID, IL, KY, ME, MI, MO, NY, OR, SC, WY 14

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005, 2007, and 2009 Reading and Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Figure 16. Example of discrepancies between NAEP and state measures of change in achievement



To compare NAEP and state changes in achievement from 2007 to 2009, we compute the difference between (a) the percentage of students reported to be meeting the state standard in 2009 and (b) the percentage of the NAEP students in 2009 that is above the NAEP scale equivalent of the state standard in 2007. Figure 16 illustrates, using hypothetical data, how the discrepancies between NAEP and state measures of change in achievement are determined.

- In State A, 85.4 percent of the students met the state's standard in 2007. This matches the percentage meeting the NAEP equivalent of the 2007 standard in 2007, by definition.
- In the top chart of the display, 91.6 percent of the students in 2009 met State A's 2007 standard, while 97.4 percent met the NAEP equivalent of the 2007 state standard in 2009.
- The change in achievement measured by the state test is 6.2 percentage points and the change in achievement measured by NAEP is 12 percentage points.
- The discrepancy between gains reported by the state and by NAEP is, therefore, 5.8 percentage points (12.0 – 6.2 = 5.8). NAEP reports larger gains than the state.
- This discrepancy is equivalent to the difference between (a) the percentage of NAEP students in 2009 that are above the NAEP equivalent of the 2007 state standard and (b) the percentage meeting the state standard in 2009 (97.4 – 91.6 = 5.8 percentage points), since the difference between the 2007 state and NAEP scores is zero by definition.
- A positive significant value for the discrepancy *D* indicates that NAEP results show more positive changes (e.g., larger gains or smaller losses) than state results. Conversely, a negative significant value indicates that state results show more positive changes than NAEP results. In the example at the bottom chart of the display, the state shows larger gains than those measured by the mapping. A non-significant value for *D* indicates that the two assessments are measuring equivalent changes in student achievement.

A more detailed discussion about comparing changes in achievement is available in the Technical Notes.

Reading

In both periods, 2005 to 2009 and 2007 to 2009, states reported more positive changes on their state reading assessment when compared with the changes measured by NAEP, with the exception of grade 4 reading from 2005 to 2009, when most states did not show significant differences between NAEP and state assessment changes in achievement.

Of the 40 states with comparable data between 2007 and 2009, the results of 22 states' assessments showed more positive changes in grade 4 reading compared with NAEP assessments, 4 states showed less positive change, and 14 states showed an equivalent change. In grade 8 reading, of 40 states, 20 states showed more positive changes from 2007 to 2009 compared with NAEP, 3 showed a less positive change, and 17 were similar. Figure 17 groups the states according to how their change in the percentages of students meeting the state standard from 2007 to 2009 compare with the changes in the percentages of students meeting the NAEP equivalent of the same state standard in the same period.

Figure 18 groups the states by how their assessment gains compared with NAEP gains from 2005 to 2009. During this period, in grade 4 reading, 10 of 17 states with comparable data showed an equivalent change on the two assessments, 4 states showed a more positive change than NAEP, and 3 showed a less positive change. In grade 8 reading, 12 out of 18 states' assessments showed more positive changes from 2005 to 2009 compared with NAEP.

Figure 17. States according to how their changes in reading achievement compared with NAEP's for the same period, by grade: 2007 to 2009

Comparison result	Grade 4	Grade 8
No difference (D=0)	AK, AL, CO, KY, LA, MA, MD, ND, NH, NM, RI, TX, UT, VT 14	AK, AL, CO, CT, FL, KY, MI, NH, NM, NV, OR, PA, RI, TN, UT, WA, WI 17
NAEP results show more positive changes than state results (D > 0)	MI, MO, WA, WY 4	ND, OH, WY 3
State results show more positive changes than NAEP results (D < 0)	AR, AZ, CA, CT, DE, FL, GA, HI, IA, ID, KS, ME, MN, MT, NV, NY, OH, OR, PA, TN, VA, WI 22	AR, AZ, CA, DE, GA, HI, IA, ID, KS, LA, MA, MD, ME, MN, MO, MT, NY, TX, VA, VT 20
Total number of states	40	40

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 and 2009 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Figure 18. States according to how their changes in reading achievement compared with NAEP's for the same period, by grade: 2005 to 2009

Comparison result	Grade 4	Grade 8
No difference (D=0)	AK, CO, IA, MA, MD, ND, NM, TN, TX, WI 10	AK, CO, IA, ND 4
NAEP results show more positive changes than state results (D > 0)	AL, FL, WA 3	OH, WI 2
State results show more positive changes than NAEP results (D < 0)	AR, CA, LA, OH 4	AL, AR, AZ, CA, FL, LA, MD, NM, NV, PA, TN, TX 12
Total number of states	17	18

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2009 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Mathematics

In mathematics, 41 states in grade 4 and 40 in grade 8 had state assessments that were comparable for 2007 and 2009. Figure 19 displays the states by whether they showed different changes in achievement between 2007 and 2009 compared with NAEP. In grade 4, 21 states showed a more positive change from 2007 to 2009 in their mathematics assessment compared with NAEP, 3 showed a less positive change, and 17 showed changes in achievement in their own test that are corroborated by NAEP results. In grade 8, 17 state assessments showed a positive change compared with NAEP, 5 showed a less positive change, and 18 had comparable changes.

Figure 20 displays the states by how their mathematics assessment compared with NAEP in terms of achievement change from 2005 to 2009. Of 16 states in the grade 4 analysis sample, 8 states showed more positive change on their assessment compared with the change based on their NAEP equivalent score. In grade 8, 6 of the 16 state assessments showed more positive change in achievement than NAEP, and 10 states had comparable changes in the sense that state assessment and NAEP measures of changes in percentages of students meeting the state standards are not statistically significantly different from each other.

Figure 19. States according to how their changes in mathematics achievement compared with NAEP's for the same period, by grade: 2007 to 2009

Comparison result	Grade 4	Grade 8
No difference (D=0)	AK, AL, AZ, CO, HI, IA, KS, MA, MD, ME, MO, MT, ND, NV, SD, TN, UT 17	AZ, CO, CT, FL, IA, MA, ME, MN, MO, ND, NH, OH, PA, SD, VT, WA, WI, WY 18
NAEP results show more positive changes than state results (D > 0)	NM, WA, WY 3	AK, MT, NV, OR, UT 5
State results show more positive changes than NAEP results (D < 0)	AR, CA, CT, DE, FL, ID, KY, LA, MI, MN, NC, NH, NY, OH, OR, PA, RI, TX, VA, VT, WI 21	AL, AR, DE, HI, ID, KS, KY, LA, MD, MI, NC, NM, NY, RI, TN, TX, VA 17
Total number of states	41	40

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 and 2009 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Figure 20. States according to how their changes in mathematics achievement compared with NAEP's for the same period, by grade: 2005 to 2009

Comparison result	Grade 4	Grade 8
No difference (D=0)	AL, CO, IA, LA, MA, ND 6	AK, AZ, CO, IA, LA, ND, NV, PA, TN, WI 10
NAEP results show more positive changes than state results (D > 0)	NM, WA 2	—
State results show more positive changes than NAEP results (D < 0)	AK, AR, CA, FL, MD, TN, TX, WI 8	AR, FL, MD, NM, OH, TX 6
Total number of states	16	16

— No state where NAEP results showed larger gains or smaller losses than state results.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2009 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Conclusion

Mapping state standards for proficient performance on the NAEP scales showed wide variation among states in the rigor of their standards. The implication is that students of similar academic skills but residing in different states are being evaluated against different standards for proficiency in reading and mathematics. All NAEP scale equivalents of states' reading standards were below NAEP's *Proficient* range; in mathematics, only one state's NAEP scale equivalent was in the NAEP *Proficient* range (Massachusetts in grades 4 and 8). In many cases, the NAEP scale equivalent for a state's standard, especially in grade 4 reading, mapped below the NAEP achievement level for *Basic* performance. There may well be valid reasons for state standards to fall below NAEP's *Proficient* range. The comparisons simply provide a context for describing the rigor of performance standards that states across the country have adopted.

Between 2007 and 2009, about one-fifth of the states changed aspects of their assessment policies or the assessment itself to the extent that their reading or mathematics results are not comparable across these two years.

Either explicitly or implicitly, such states adopted new performance standards. By mapping the state standards in both years to the same NAEP scale, the changes in rigor of the standards can be measured. When examined across grades 4 and 8 for both reading and mathematics, of the 34 instances where the states reported changes in their assessments, the rigor of the standards increased in 21 of them, did not change in 8, and decreased in 5 as measured by NAEP scale equivalents.

The remaining states made no changes to their assessment policies or made changes that were minor enough that their test results remained comparable. In more than half of the 40 states that indicated no substantive changes in their state reading assessments (24 states in grade 4 and 21 states in grade 8), the differences between their 2007 and 2009 NAEP equivalent scores were statistically significant. In most cases, the 2009 scores were lower (22 out of 24 states in grade 4 and 19 out of 21 states in grade 8). In mathematics, in the majority of the states with no substantive changes in their state assessments, the differences between their

2007 and 2009 NAEP equivalent scores were not statistically significant (21 out of 41 states in grade 4 and 22 of 40 states in grade 8). However, in 17 of 41 states in grade 4 and in 16 of 40 states in grade 8, the 2009 NAEP scale equivalents of state standards were lower.

For the same groups of states (i.e., states whose assessments did not change), it was possible to check the extent to which NAEP corroborated the changes in achievement measured in the states' assessments between 2007 and 2009. In both subjects, NAEP's measurements of student progress did not agree with the progress measured by state assessment in at least half the states. In most cases, states' results showed larger gains or smaller losses than did NAEP. These findings of disagreements between the two measures could be explained by a methodological change in one of the tests (e.g., accommodations, scaling, time of administration, or exclusions) or by differences between NAEP and the state test domains affecting the skills learned by students and tested in the two assessments.

In this report we conducted three sets of analyses—assessing the relative rigor of state standards, describing changes in relative rigor of standards when states establish new policies or testing systems, and corroborating state progress in student performance—the results of which show that NAEP, as a common yardstick, continues to be an essential benchmark for states in evaluating their standards.

References

Bandeira de Mello, V., Blankenship, C., and McLaughlin, D.H. (2009). *Mapping State Proficiency Standards Onto NAEP Scales: 2005–2007* (NCES 2010-456). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

McLaughlin, D.H. (2008). *Measurement Error in Comparing NAEP and State Test Gains.* Palo Alto, CA: American Institutes for Research.

McLaughlin, D.H., Bandeira de Mello, V., Blankenship, C., Chaney, K., Esra, P., Hikawa, H., Rojas, D., William, P., and Wolman, M. (2008a). *Comparison Between NAEP and State Reading Assessment Results: 2003* (NCES 2008-474). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

McLaughlin, D.H., Bandeira de Mello, V., Blankenship, C., Chaney, K., Esra, P., Hikawa, H., Rojas, D., William, P., and Wolman, M. (2008b). *Comparison Between NAEP and State Mathematics Assessment Results: 2003* (NCES 2008-475). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

U.S. Department of Education, National Center for Education Statistics (2007). *Mapping 2005 State Proficiency Standards Onto the NAEP Scales* (NCES 2007-482). Washington, DC: Author. Retrieved July 14, 2007, from http://nces.ed.gov/nationsreportcard/pdf/studies/2007482.pdf.

Technical Notes

NAEP Achievement Levels

NAEP uses both scale scores and achievement levels to report student performance. Scale scores show what students know and can do, and achievement levels are performance standards for what students should know and be able to do. The NAEP achievement levels *Basic*, *Proficient*, and *Advanced* are used to interpret the meaning of the NAEP scales. They are indicators of student performance. *Basic* denotes partial mastery of the knowledge and skills that are fundamental to proficient work at a given grade. *Proficient* represents solid academic performance. Students reaching this level have demonstrated competency on challenging subject matter. However, *Proficient* is not synonymous with grade-level performance. *Advanced* signifies superior performance. These achievement levels are set independently by the National Assessment Governing Board, which sets policy for NAEP.

NCES has determined (as provided by NAEP's authorizing legislation) that NAEP achievement levels should continue to be used on a trial basis and should be interpreted with caution (see http://nces.ed.gov/nationsreportcard/achlevdev.asp?).

Estimation Methods

Estimation of the placement of state performance standards on the NAEP scale

This section summarizes the estimation methods used in the mapping procedure to place state performance standards onto the NAEP scales. The following description of the method is excerpted from the 2009 mapping report available at http://nces.ed.gov/nationsreportcard/pdf/studies/2010456.pdf.

The method of obtaining *equipercentile equivalents* involves the following steps:

a. Obtain for each school in the NAEP sample the proportion of students in that school who meet the state performance standard on the state's test.

b. Estimate the state proportion of students who meet the standard on the state test, by weighting the proportions (from step 1) for the NAEP schools, using NAEP school weights.

c. Estimate the weighted distribution of scores on the NAEP assessment for the state as a whole, based on the NAEP sample of schools and students within schools.

d. Find the point on the NAEP scale at which the estimated proportion of students in the state who score above that point (using the distribution obtained in step 3) equals the proportion of students in the state who meet the state's own performance standard (obtained in step 2).

The reported percentage meeting the state's standard in each NAEP school s, p_s, is used to compute a state percentage meeting the state's standards, p_S, using the NAEP school weights, w_s. For each school, w_s is the sum of the student weights, w_{is}, for the students selected for NAEP in that school.[1] For each of the five sets of NAEP plausible values, v = 1 through 5, we solve the following equation for c, the point on the NAEP scale corresponding to the percentage meeting the state's standard:[2]

$$p_S = \sum_{is, s \in S} w_{is} p_s \Big/ \sum_{is, s \in S} w_{is} \qquad [1]$$

$$= \sum_{is, s \in S} w_{is} \partial_{isv}(c) \Big/ \sum_{is, s \in S} w_{is} \qquad [2]$$

where the sum is over students in schools participating in NAEP, and $\partial_{isv}(c)$ is an indicator variable that is 1 if the v-th plausible value for student i in school s, y_{isv}, is greater than or equal to c, and 0 otherwise. The five values of c obtained for the five sets of plausible values are averaged to produce the NAEP threshold corresponding to the state standard, that is, the reported mapping of the standard onto the NAEP scale. Variation in results over the five sets of plausible values is a component of the standard error of the estimate, which is computed by following standard NAEP procedures.

An estimate of the standard error of the mapping is necessary to test the question of whether the NAEP scale equivalent of the standard is stable across the two years. If we denote the NAEP scale equivalent of the standard in year Y by $\hat{c}_Y$, then the standard error of the difference, $\hat{c} = \hat{c}_1 - \hat{c}_2$, is just the square root of the sum of the squares of the standard errors of the two separate NAEP scale equivalents. That is, $SE(\hat{c}) = \sqrt{SE(\hat{c}_1)^2 + SE(\hat{c}_2)^2}$.

Each can be estimated by applying the NAEP jackknife technique to the mapping process.

Relative error

When used to place state standards on the NAEP scale, equipercentile mapping will produce an answer even if NAEP and state assessment scores are completely unrelated to each other. Some additional data, beyond the percentage meeting the standard in the state and the distribution of NAEP plausible values—the only data used in the computation—are needed to test the validity of the mapping.

To evaluate the validity of the placement of a state standard on the NAEP scale, we measure how well the procedure reproduces the percentages reported by the state as meeting the standard in each NAEP-participating school. If the mapping is valid, the procedure should reproduce the individual school percentages fairly accurately. However, if the state assessment and NAEP are measuring different, uncorrelated characteristics of students, the school-level percentages meeting the state standard as measured by NAEP will bear no relationship to the school-level percentages meeting the state's standards as reported by the state.

The correlation coefficient showing the relationship between the percentages reported for schools by the state and those estimated from the NAEP scale equivalents provides a straightforward measure of the appropriateness of the mapping. However, it does not indicate the amount of error that is added to the placement of the standard by the fact that NAEP and the state assessment may not measure the same construct. We must determine how high the correlation must be to justify inferences that are based on the mapping. Also needed is a measure of that error, as a fraction of the total variation of percentages meeting the standard across schools.

The NAEP estimate of the percentage meeting the standard in a school is subject to both sampling and measurement error. However, even if the NAEP measure had no sampling or measurement error, and even if NAEP measured exactly the same construct as the state assessment, NAEP would not reproduce exactly the state assessment percentage for each school. The difference occurs because the state assessment scores are based on different administrations, at different times of year, with different motivational contexts and different rules for exclusion and accommodation. The state assessment scores are also subject to measurement error, although for school-level aggregates, the measurement error is smaller than it is for individual student estimates.

Although we recognize that discrepancies between the reported figure from each school and the estimate based on the NAEP mapping will occur, it is, nevertheless, important that the discrepancies be small relative to the variation in outcomes across schools. If the variance of the discrepancies is more than a fraction of the total variance across schools in percentage meeting a standard, the validity of the placement of the standard could be considered suspect, even though the nominal standard error of the state-level estimate may be small.

To evaluate the mapping, we therefore compare three variances:

1. total variance of reported percentages meeting the state's standard across the schools participating in NAEP in the state, $\sigma^2(p_s)$;

2. average squared deviation between the reported percentage, p_s and the percentage based on the NAEP mapping for each school s, $\hat{p}_s$: $average_s\ (p_s - \hat{p}_s)^2$; and

3. average expected sampling and measurement error in the NAEP estimate for each school s, $average_s\ (\hat{p}_s - \mathrm{E}(\hat{p}_s))^2$.

We estimate the sizes of what the (squared) discrepancies would have been if NAEP were not subject to sampling and measurement error by subtracting quantity (3) from quantity (2), and we compare these adjusted (squared) discrepancies with the overall variation in percentages across schools $\sigma^2(p_s)$ (quantity (1)). If the adjusted (squared) discrepancies correspond to a large component the overall variance of the percentages, the NAEP data do not reproduce the school-level percentages with sufficient accuracy to justify inferences based on the placement of the standard on the NAEP scale. That is, we want the relative error $K < k$,

$$K = \left[\left(average_s\ (p_s - \hat{p}_s)^2 - average_s\ (\hat{p}_s - \mathrm{E}(\hat{p}_s))^2\right) \Big/ \sigma^2(p_s)\right] < k \quad [3]$$

where $0 \le k \le 1$.

We want the discrepancy variance (*2*) to be less than a threshold k of the variance in the state test score school percentages (*1*), but we do not want to penalize the mapping for the measurement and sampling error in $\hat{p}_s$ (quantity *3*), which contributes to quantity (*2*). Therefore, we subtract (*3*) from (*2*) before dividing by (*1*). The resulting numerator of the relative error K is an estimate of the amount of discrepancy variance that cannot be accounted for by NAEP sampling and measurement error. Because both quantities (*2*) and (*3*) are sample estimates of variances, it is reasonable to expect that they will usually differ from the true variances of (*2*) and (*3*), and this can lead to (*2*) – (*3*) < 0 in some cases. In fact, if there were no linking error, we would expect (*2*) – (*3*) < 0 in half the cases, because (*2*) and (*3*) would be two estimates of the same variance.

Both the discrepancies and the estimation of NAEP random estimation error are more stable in schools with larger NAEP samples of students. Therefore, to increase the stability of the estimate of *K*, the average over schools was weighted according to the size of the NAEP sample of students in the school; a small number of NAEP schools with fewer than five NAEP participants are not included in the computations.

The NAEP random estimation error variance is the sum of two components, sampling error and measurement error. Because at the student level the variable of interest is a simple binomial variable (meets or does not meet the standard), to estimate the sampling variance we can use the binomial variance of the estimate of a percentage, $\hat{p}_s(100 - \hat{p}_s)/n_s$, where n_s is the size of the NAEP sample in the school and $\hat{p}_s$ is the percentage of NAEP participants in the school with plausible values greater than the value estimated to be equivalent to the state standard. The binomial variance should be reduced by a finite population correction, $fpc = \sqrt{(N_s - n_s)/(N_s - 1)}$, because the NAEP sample is a sizeable fraction of the number of students in the particular grade, N_s, at most schools. If the number of students per grade is not known, the average finite population correction for schools with NAEP samples of the same size is used.

NAEP measurement error is estimated by the variance of the five estimates for each school's percentage meeting the standard, based on the five alternative sets of plausible values v, for the participating students, $\sigma_v^2(\hat{p}_{s,v})$. Because $\hat{p}_s$ is computed as the average of values based on five plausible value sets, the measurement error component is divided by 5. Thus, the quantity in (*3*) above is estimated by

$$\mathrm{E}\,(\hat{p}_s - \mathrm{E}(\hat{p}_s))^2 = (p_s q_s / n_s)(fpc)^2 + \sigma_v^2(\hat{p}_{s,v})/5. \quad [4]$$

In this study, the criterion proposed is to consider relative errors greater than .5 as indicating that the mapping error is too large to support any useful inferences from the placement of the standard on the NAEP scale.

Setting the criterion for the validity of this application of the equipercentile mapping method at K = .5 is arbitrary but plausible. Clearly, it should not be taken as an absolute inference of validity—two assessments, one with a relative error of .6 and

the other with .4, have similar validity. Setting a criterion serves to call attention to the cases in which we should consider a limitation on the validity of the mapping as an explanation for otherwise unexplainable results. Although estimates of standards with greater relative error because of differences in measures are not thereby invalidated, any inferences based on them require additional evidence. For example, a finding of differences in trend measurement between NAEP and a state assessment when the standard mapping has large relative error may be explainable in terms of unspecifiable differences between the assessments, ruling out further comparison. Nevertheless, because the relative error criterion is arbitrary, results for all states are included in the report and in the discussion of findings, irrespective of the relative error of the mapping of the standards.

Notes

1. To ensure that NAEP and state assessments are equitably matched, NAEP schools that are missing state assessment scores (i.e., small schools, typically representing approximately 4 percent of the students in a state) are excluded from this process. Even if the small excluded schools perform differently from included schools, no substantial bias in the estimation process would be introduced, unless their higher or lower scoring was specific to NAEP or specific to the state assessment.

2. Estimations of NAEP scale score distributions are based on an estimated distribution of possible scale scores (or *plausible values*), rather than point estimates of a single scale score. More details are available at http://nces.ed.gov/nationsreportcard/tdw/analysis/est_pv_individual.asp.

Comparing NAEP and State Measures of Change

When state and NAEP assessments remain the same over two assessment periods, NAEP can be used to corroborate progress on the state assessments. If either NAEP or a state test has substantively changed between the two years, then comparisons of achievement changes identified by the two tests cannot be justified.

To compare NAEP and state changes in achievement from 2007 to 2009, we compute the difference between (a) the percentage of students reported to be meeting the state standard in 2009 and (b) the percentage of the NAEP students in 2009 that is above the NAEP scale equivalent of the state standard in 2007.

Computing the discrepancies between NAEP and state measures of changes in achievement

Let D be the discrepancy between NAEP and state changes in achievement from year 1 to year 2.

$$D = (D_N - D_S)$$

where D_S is the change from year 1 to year 2 in achievement measured by the state test, and D_N is the change from year 1 to year 2 in achievement measured by the mapping.

The change D_N is

$$D_N = (P_{2N} - P_{1N})$$

where P_{2N} is the percentage of the NAEP students in year 2 that are above the NAEP scale equivalent of the state standard in year 1, and P_{1N} is the percentage of the NAEP students in year 1 that are above the NAEP scale equivalent of the state standard in year 1.

Similarly, the change D_S is

$$D_S = (P_{2S} - P_{1S})$$

where P_{2S} is the percentage of students reported to be meeting the state standard in year 2, and P_{1S} is the percentage of students reported to be meeting the state standard in year 1.

For the year for which the NAEP scale equivalent is computed, the percentage meeting the state's standard and the percentage meeting the NAEP scale equivalent are, by definition, the same.

$$P_{1S} = P_{1N}$$

Therefore, the discrepancy D is the difference between (a) the percentage of students reported to be meeting the state standard in year 2 and (b) the percentage of the NAEP students in year 2 that are above the NAEP scale equivalent of the state standard in year 1.

$$D = (P_{2N} - P_{1N}) - (P_{2S} - P_{1S})$$

$$D = (P_{2N} - P_{2S}) - (P_{1N} - P_{1S})$$

$$D = (P_{2N} - P_{2S})$$

When $D > 0$ (i.e., $D_N > D_S$ or equivalently, $P_{2N} > P_{2S}$) the change from year 1 to year 2 measured by the mapping is more positive (or less negative) than the change from year 1 to year 2 measured by the state test. For $D < 0$, that is ($D_N < D_S$ or equivalently, $P_{2N} < P_{2S}$), the change measured by the mapping is less positive (or more negative) than the change measured by the state test.

The expectation is that both the state assessments and NAEP would show the same changes in achievement between the two years. Statistically significant differences between NAEP and state measures of changes in achievement indicate that more progress is made on either the NAEP skill domain or the state-specific skill domain between two years. A more positive change on the state test (larger gains or smaller losses) indicates that students gained more on the state-specific skill domain. For example, a focus in instruction on state-specific content might lead a state assessment to show more progress in achievement than NAEP. Similarly, a less positive change on the state test indicates that students gained more on the NAEP skill domain. For example, a focus in instruction on NAEP content that is not a part of the state assessment might lead the state assessment to show progress in achievement that is less than that of NAEP.

To measure achievement changes in terms of percentages of students meeting a standard requires that the standards remain unchanged. If the standards have changed, one cannot be certain whether achievement gains are due to gains in achievement or to a lowering of the standard, for example. Similarly, if one observes a loss in achievement and the standards have changed, one cannot be certain if it is due to a real achievement loss or an increase in the standards. Therefore, when both NAEP and a state's standard remain unchanged between two years, the question of whether NAEP and the state assessment agree on the size of an achievement change is the same as the question of whether the mapping of the state's standard onto the NAEP scale is stable over the two years.

Measuring the standard error of D

Because the data available for mapping states' standards onto the NAEP scale are limited to school-level percentages of students achieving a state's standard in schools participating in NAEP, the critical statistic for comparing NAEP versus state-test score changes is

$$D = (\hat{p}_{2N|map=1} - \hat{p}_{2S}) - (\hat{p}_{1N|map=1} - \hat{p}_{1S}) \qquad [5]$$

where $\hat{p}_{YS}$ is the state percentage meeting the standard in year Y, estimated by the weighted average of the percentages in the NAEP schools, and $\hat{p}_{YN|map=1}$ is the percentage of the distribution of NAEP plausible values in the state in year Y, estimated by the (same) weighted average of the distributions in the NAEP schools, which are above the NAEP scale value that was found in year 1 to correspond to the state standard.

For example, if the state shows a gain from 50 percent to 60 percent meeting the standard and NAEP reports a gain from 50 percent to 55 percent meeting the state's standard, then D = (55 – 60) – (50 – 50) = -5. The statistical question to be addressed is whether a value of 5 for D is larger than we would expect on the basis of measurement and sampling error.

The term in the second parenthesis of equation [5] is zero by definition, with no error, because the NAEP scale value onto which the state's standard is mapped (in year 1) is the value that forces an exact match of percentages (in year 1). That is not to say that $\hat{p}_{1S}$ and $\hat{p}_{1N|map=1}$ are error-free estimates of their respective population statistics, just that the second term in D is exactly zero. The errors in $\hat{p}_{1S}$ and $\hat{p}_{1N|map=1}$ contribute to the error in the other term $(\hat{p}_{2N|map=1} - \hat{p}_{2S})$ through mapping error.

Both NAEP estimates, $\hat{p}_{1N|map=1}$ and $\hat{p}_{2N|map=1}$, are based on percentages of the student score distribution meeting the same scale value, the one mapped from the year 1 data. To measure achievement changes in terms of percentages of students meeting a standard, it is necessary to use exactly the same standard for both years.[3] In fact, if achievement changes are measured purely in terms of percentages meeting a standard, finding an achievement gain in the population is equivalent to finding that the test became easier for the population to meet the standard. In other words, unless we are assured that the standard has not been lowered, we cannot infer that finding that the standard became easier for the population means that the population's achievement increased. We cannot exclude the possibility that the standard was lowered unless we have evidence to exclude it. An example of that evidence is finding that in both years, the standard is equivalent to the same NAEP score, if we assume that NAEP remained unchanged between the years. Thus, the question of whether NAEP and the state assessment agree on the size of achievement change is virtually equivalent to the question of whether the mapping of the state's standard onto the NAEP scale was stable over the two years.

Because the second term in the equation for D is zero, we can redefine D as

$$D = (\hat{p}_{2N|map=1} - \hat{p}_{2S}) \quad [6]$$

and focus on the estimation of the sources of error; that is, on the expected variation between D and the value it would take on if the estimates of the percentages meeting the standard were equal to their population values, $\hat{p}_{2S}$ and $\hat{p}_{2N|map=1}$.

Many factors contribute to random variation of D around its true value, which would be zero if NAEP and the state assessments show the same gains/losses.[4] However, in view of the complexity of any psychometric model for D, the most robust procedure for estimating the standard error of D is the standard NAEP procedure, combining NAEP measurement error, estimated by variation in values of D obtained for each of the five plausible value sets, with NAEP sampling error, estimated by the NAEP jackknife technique.

Additional information on comparing NAEP and state measures of change is available at http://nces.ed.gov/nationsreportcard/pdf/studies/2010456.pdf.

Notes

3. If we were to estimate $\hat{p}_{2N}$ from a mapping based on year 2 data, D would be identically zero, a meaningless result.

4. These factors are discussed in McLaughlin (2008).

Appendix Tables

Table A-1. NAEP scale equivalent scores for state reading proficiency standards at grades 4 and 8 in 2009, and their differences from the 2005 and 2007 estimates of the same standards, by state

	Reading Grade 4					Reading Grade 8				
State	2009 NAEP scale equivalent	Change from 2007 NAEP scale equivalent	Change from 2005 NAEP scale equivalent	2007 and 2009 tests comparable	2005 and 2009 tests comparable	2009 NAEP scale equivalent	Change from 2007 NAEP scale equivalent	Change from 2005 NAEP scale equivalent	2007 and 2009 tests comparable	2005 and 2009 tests comparable
Alabama	179	#	7*	√	√	234	#	-3	√	√
Alaska	183	-1	1	√	√	231	-2	1	√	√
Arizona	193	-5*	—	√	√	241	-4*	-3	√	√
Arkansas	200	-13*	-17*	√	√	241	-8*	-13*	√	√
California	202	-8*	-8*	√	√	259	-3*	-4*	√	√
Colorado	183	-4	-3	√	√	228	-2	#	√	√
Connecticut	208	-5*	-4*	√		243	-2	1	√	
Delaware	199	-4*	—	√		236	-3*	-6*	√	
District of Columbia	205	—	—	√		244	—	—	√	
Florida	206	-3*	4*	√	√	262	#	-3*	√	√
Georgia	178	-7*	4	√		209	-7*	-15*	√	
Hawaii	203	-9*	-2	√		241	-3*	-20*	√	
Idaho	186	-11*	1	√		218	-14*	-17*	√	
Illinois	198	-1	—			234	-2	-11*		
Indiana	203	4*	4*			255	4*	5*		
Iowa	194	-5*	-3	√	√	248	-4*	-2	√	√
Kansas	186	-6*	—	√		236	-5*	-6*	√	
Kentucky	205	#	-1	√		253	2	—	√	
Louisiana	192	-1	-5*	√	√	243	-3	-8*	√	√
Maine	207	-6*	-17*	√		253	-8*	-23*	√	
Maryland	187	1	#	√	√	237	-13*	-8*	√	√
Massachusetts	234	2	#	√	√	249	-3*	—	√	
Michigan	194	16*	12*	√		236	-2	—	√	
Minnesota	204	-11*	—	√		259	-6*	—	√	
Mississippi	210	46*	49*			254	3*	8*		
Missouri	229	2	—	√		267	-5*	—	√	
Montana	198	-5*	1	√		246	-4*	-7*	√	
Nebraska	—	—	—	√		—	—	—	√	
Nevada	202	-5*	—	√		246	-2	-7*	√	√
New Hampshire	211	1	—	√		256	-2	—	√	
New Jersey	221	20*	31*			244	-8*	-6*		
New Mexico	207	-3*	-1	√	√	246	-2	-5*	√	√
New York	200	-9*	-7*	√		247	-13*	-21*	√	
North Carolina	204	22*	21*			246	29*	30*		
North Dakota	203	1	-1	√	√	253	2	-2	√	√
Ohio	192	-6*	-7*	√	√	251	12*	11*	√	√
Oklahoma	211	40*	29*			249	17*	5*		
Oregon	177	-8*	—	√		250	-1	-4*	√	
Pennsylvania	206	-6*	—	√		245	#	-13*	√	√
Rhode Island	209	-1	—	√	√	252	-2	—	√	√
South Carolina	194	-29*	-35*			245	-36*	-32*		
South Dakota	199	13*	—			254	5*	—		
Tennessee	170	-4*	1	√	√	211	#	-11*	√	√
Texas	188	1	-2	√	√	201	-21*	-24*	√	√
Utah	196	-1	—	√	√	235	1	—	√	
Vermont	214	#	—	√	√	259	-5*	—	√	√
Virginia	186	-5*	—	√		229	-10*	-14*	√	
Washington	205	3	8*	√	√	253	#	—	√	
West Virginia	206	24*	20*			249	20*	21*		
Wisconsin	189	-4	#	√	√	232	2	3	√	√
Wyoming	208	4*	-20*	√		259	12*	-19*	√	

— Not available; # Rounds to zero; * Statistically different from zero ($p < .05$); √ State assessment is comparable between years when state confirmed making no substantive changes in the assessment;

NOTE: Blank cell indicates state assessment is not comparable between years.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005, 2007, and 2009 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Table A-2. NAEP scale equivalent scores for state mathematics proficiency standards at grades 4 and 8 in 2009, and their differences from the 2005 and 2007 estimates of the same standards, by state

	Mathematics Grade 4					Mathematics Grade 8				
State	2009 NAEP scale equivalent	Change from 2007 NAEP scale equivalent	Change from 2005 NAEP scale equivalent	2007 and 2009 tests comparable	2005 and 2009 tests comparable	2009 NAEP scale equivalent	Change from 2007 NAEP scale equivalent	Change from 2005 NAEP scale equivalent	2007 and 2009 tests comparable	2005 and 2009 tests comparable
Alabama	207	1	#	√	√	246	-7*	—	√	√
Alaska	218	1	-4*	√	√	268	3	#	√	√
Arizona	212	-1	—	√	√	266	-2	1	√	√
Arkansas	216	-13*	-20*	√	√	267	-9*	-20	√	√
California	220	-5*	-10*	√	√	—	—	—		
Colorado	202	1	1	√	√	256	-3	-2	√	√
Connecticut	214	-6*	-7*	√		251	-1	-6*	√	
Delaware	220	-5*	—	√		269	-3*	-7*	√	
District of Columbia	217	—	—	√		258	—	—	√	
Florida	225	-5*	-6*	√	√	266	-1	-4*	√	√
Georgia	218	5*	3			247	4	-8*		
Hawaii	239	1	-8*	√		286	-8*	-10*	√	
Idaho	213	-5*	6	√		261	-3*	-4*	√	
Illinois	207	-1	—			251	#	-25*		
Indiana	229	2	4*			273	7*	7*		
Iowa	221	1	2	√	√	263	-1	1	√	√
Kansas	217	-2	-1	√		265	-5*	—	√	
Kentucky	223	-6*	—	√		273	-6*	-12*	√	
Louisiana	221	-2	-2	√	√	263	-4*	-1	√	√
Maine	234	-2	-14*	√		284	-2	-15*	√	
Maryland	208	1	-7*	√	√	271	-7*	-5*	√	√
Massachusetts	255	1	#	√	√	300	-2	-1	√	
Michigan	200	-4*	-22*	√		253	-7*	-16*	√	
Minnesota	233	-5*	—	√		287	1	—	√	
Mississippi	223	19*	17*			264	1	2		
Missouri	246	1	3*	√		287	-2	-24*	√	
Montana	235	1	14*	√		285	3	14*	√	
Nebraska	—	—	—	√		—	—	—	√	
Nevada	225	2	—	√		269	2	-1	√	√
New Hampshire	237	-2	—	√		281	-1	—	√	
New Jersey	231	11*	10*			272	#	-1		
New Mexico	236	4*	4*	√	√	277	-8*	-10*	√	√
New York	207	-12*	#	√		249	-24*	-26*	√	
North Carolina	220	-11*	18*	√		253	-17*	6*	√	
North Dakota	225	-1	1	√	√	278	-1	1	√	√
Ohio	219	-5*	-13*	√		265	#	-9*	√	√
Oklahoma	228	15*	10*			269	20*	11*		
Oregon	214	-6*	—	√		266	3*	-3*	√	
Pennsylvania	218	-5*	—	√		272	1	0	√	√
Rhode Island	231	-4*	—	√	√	275	-4*	—	√	√
South Carolina	215	-30*	-31*			270	-42*	-36*		
South Dakota	224	#	—	√	√	271	1	—	√	√
Tennessee	195	-3	-4*	√	√	229	-5	-1	√	√
Texas	214	-3	-5*	√	√	254	-14*	-18*	√	√
Utah	225	2	—	√	√	275	19*	—	√	√
Vermont	236	-3*	—	√	√	282	-1	—	√	√
Virginia	213	-6*	—	√		251	-8*	-2	√	
Washington	243	4*	8*	√	√	288	2	—	√	
West Virginia	225	9*	11*			270	16*	17*		
Wisconsin	219	-4	-6*	√	√	262	#	-2	√	√
Wyoming	226	9*	-25*	√		278	-2	-15*	√	

— Not available; # Rounds to zero; * Statistically different from zero ($p < .05$); √ State assessment is comparable between years when state confirmed making no substantive changes in the assessment;

NOTE: Blank cell indicates state assessment is not comparable between years.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005, 2007, and 2009 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Table A-3. Direction of change in the estimated NAEP scale equivalent scores of state reading proficiency standards for the states that did not make significant changes in their assessments, by grade and comparison result: 2007 to 2009

Comparison result	Grade 4	Grade 8
Increase	MI, WY 2	OH, WY 2
No significant change	AK, AL, CO, KY, LA, MA, MD, MO, ND, NH, RI, TX, UT, VT, WA, WI 16	AK, AL, CO, CT, FL, KY, LA, MI, ND, NH, NM, NV, OR, PA, RI, TN, UT, WA, WI 19
Decrease	AR, AZ, CA, CT, DE, FL, GA, HI, IA, ID, KS, ME, MN, MT, NM, NV, NY, OH, OR, PA, TN, VA 22	AR, AZ, CA, DE, GA, HI, IA, ID, KS, MA, MD, ME, MN, MO, MT, NY, TX, VA, VT 19
Total number of states	40	40

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 and 2009 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

Table A-4. Direction of change in the estimated NAEP scale equivalent scores of state mathematics proficiency standards for the states that did not make significant changes in their assessments, by grade and comparison result: 2007 to 2009

Comparison result	Grade 4	Grade 8
Increase	NM, WA, WY 3	OR, UT 2
No significant change	AK, AL, AZ, CO, HI, IA, KS, LA, MA, MD, ME, MO, MT, ND, NH, NV, SD, TN, TX, UT, WI 21	AK, AZ, CO, CT, FL, IA, MA, ME, MN, MO, MT, ND, NH, NV, OH, PA, SD, TN, VT, WA, WI, WY 22
Decrease	AR, CA, CT, DE, FL, ID, KY, MI, MN, NC, NY, OH, OR, PA, RI, VA, VT 17	AL, AR, DE, HI, ID, KS, KY, LA, MD, MI, NC, NM, NY, RI, TX, VA 16
Total number of states	41	40

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2009 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, EDFacts SY 2008–09, Washington, DC, 2010. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2010.

This page intentionally left blank

This page intentionally left blank

Arne Duncan
Secretary
U.S. Department of Education

John Q. Easton
Director
Institute of Education Sciences

Jack Buckley
Commissioner
National Center for Education Statistics

The National Center for Education Statistics (NCES), located within the U.S. Department of Education and the Institute of Education Sciences, is the primary federal entity for collecting and analyzing data related to education.

The National Assessment of Educational Progress (NAEP) is a congressionally authorized project sponsored by the U.S. Department of Education. The Commissioner of Education Statistics is responsible by law for carrying out the NAEP project.

The National Assessment Governing Board is responsible for setting policy for NAEP, including the NAEP achievement levels.

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

This report was prepared for the National Center for Education Statistics under Contract No. ED-04-CO-00025/0019 with the American Institutes for Research. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Full results can be found at: http://nces.ed.gov/nationsreportcard/studies/statemapping/.

August 2011

Suggested Citation

Bandeira de Mello, V. (2011), *Mapping State Proficiency Standards Onto the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005–2009* (NCES 2011-458). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC: Government Printing Office.

Content Contact

Taslima Rahman
(202) 502-7316
taslima.rahman@ed.gov

The Department of Education's mission is to promote student achievement and preparation for global competitiveness by fostering education excellence and ensuring equal access.

www.ed.gov