EMBARGOED
until
Oct 29,2009
10:30am EDT

Mapping State Proficiency Standards Onto NAEP Scales: 2005-2007

NCES 2010-456
U. S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

U.S. DEPARTMENT OF EDUCATION
NCES 2010-456

Mapping State Proficiency Standards Onto NAEP Scales: 2005-2007

Research and Development Report

October 2009

Victor Bandeira de Mello
Charles Blankenship
American Institutes for Research

Don McLaughlin
Statistics and Strategies

Taslima Rahman
Project Officer
National Center for Education Statistics

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Statistics
Stuart Kerachsky
Acting Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to

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

October 2009

The NCES World Wide Web Home Page address is http://nces.ed.gov.
The NCES World Wide Web Electronic Catalog address is http://nces.ed.gov/pubsearch.

Suggested Citation
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.

For ordering information on this report, write to

U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794-1398

or call toll free 1-877-4ED-Pubs or order online at http://www.edpubs.org.

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

FOREWORD

The Research and Development (R&D) series of reports at the National Center for Education Statistics has been initiated to

- Share studies and research that are developmental in nature. The results of such studies may be revised as the work continues and additional data become available;
- Share the results of studies that are, to some extent, on the cutting edge of methodological developments. Emerging analytical approaches and new computer software development often permit new and sometimes controversial analyses to be done. By participating in frontier research, we hope to contribute to the resolution of issues and improved analysis; and
- Participate in discussions of emerging issues of interest to educational researchers, statisticians, and the federal statistical community in general.

The common theme in all three goals is that these reports present results or discussions that do not reach definitive conclusions at this point in time, either because the data are tentative, the methodology is new and developing, or the topic is one on which there are divergent views. Therefore, the techniques and inferences made from the data are tentative and subject to revision. To facilitate the process of closure on the issues, we invite comment, criticism, and alternatives to what we have done.

Such responses should be directed to

Marilyn Seastrom
Chief Statistician
Statistical Standards Program
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006-5651

EXECUTIVE SUMMARY

Since 2003, the National Center for Education Statistics (NCES) has sponsored the development of a method for mapping each state's standard for proficient performance onto a common scale—the achievement scale of the National Assessment of Educational Progress (NAEP). When states' standards are placed onto the NAEP reading or mathematics scales, the level of achievement required for proficient performance in one state can then be compared with the level of achievement required in another state. This allows one to compare the standards for proficiency across states.

The mapping procedure offers an approximate way to assess the relative rigor of the states' adequate yearly progress (AYP) standards established under the No Child Left Behind Act of 2001. Once mapped, the NAEP scale equivalent score representing the state's proficiency standards can be compared to indicate the relative rigor of those standards. The term rigor as used here does not imply a judgment about state standards. Rather, it is intended to be descriptive of state-to-state variation in the location of the state standards on a common metric.

This report presents mapping results using the 2005 and 2007 NAEP assessments in mathematics and reading for grades 4 and 8. The analyses conducted for this study addressed the following questions:

- How do states' 2007 standards for proficient performance compare with each other when mapped on the NAEP scale?
- How do the 2007 NAEP scale equivalents for state standards compare with those estimated for 2005?
- Using the 2005 NAEP scale equivalent for state standards to define a state's *proficient* level of performance on NAEP, do NAEP and that state's assessment agree on the changes in the proportion of students meeting that state's standard for proficiency from 2005 to 2007?

To address the first question, the 2007 *NAEP scale equivalent* of each state reading and mathematics proficiency standard for each grade was identified. The mapping procedure was applied to the test data of 48 states.[1] Key findings of the analysis presented in Section 3 of the report are:

- In 2007, as in 2003 and 2005, state standards for *proficient* performance in reading and mathematics (as measured on the NAEP scale) vary across states in terms of the levels of achievement required. For example, the *distance* separating the five states with the highest standards and the five states with the lowest standards in grade 4 reading was comparable to the difference between *Basic* and *Proficient* performance on NAEP.[2] The distance was as large in reading at grade 8 and as large in mathematics in both grades.

[1] Test data for the District of Columbia, Nebraska, and Utah were not available to be included in the analysis. California does not test general mathematics in grade 8.

[2] NAEP defines *Proficient* as *competency over challenging subject matter*, not grade-level performance. *Basic* is defined as partial mastery of the skills necessary for *Proficient* performance.

- In both reading and mathematics, the 29- to 30-point distance separating the five highest and the five lowest NAEP scale equivalent of state standards for *proficient* performance was nearly as large as the 35 points that represent approximately one standard deviation in student achievement on the NAEP scale.
- In grade 4 reading, 31 states set grade 4 standards for proficiency (as measured on the NAEP scale) that were lower than the cut point for *Basic* performance on NAEP (208). In grade 8 reading, 15 states set standards that were lower than the *Basic* performance on NAEP (243).
- In grade 4 mathematics, seven states set standards for proficiency (as measured on the NAEP scale) that were lower than the *Basic* performance on NAEP (214). In grade 8 mathematics, eight states set standards that were lower than the *Basic* performance on NAEP (262).
- Most of the variation (approximately 70 percent) from state to state in the percentage of students scoring proficient or above on state tests can be explained by the variation in the level of difficulty of state standards for proficient performance. States with higher standards (as measured on the NAEP scale) had *fewer* students scoring proficient on state tests.
- The rigor of the state standards is not consistently associated with higher performance on NAEP. This association is measured by the squared correlation between the NAEP scale equivalent of the state standards and the percentages of students who scored at or above the NAEP *Proficient* level. In grade 4 reading and mathematics, the squared correlations are around .10 and statistically significant. In grade 8 reading and mathematics, the squared correlations are less than .07 and are not statistically significant.

To address the second question, the analyses focused on the consistency of mapping outcomes over time using both 2005 and 2007 assessments. Although NAEP did not change between 2005 and 2007, some states made changes in their state assessments in the same period, changes substantial enough that states indicated that their 2005 scores were not comparable to their 2007 scores. Other states indicated that their scores for those years are comparable. Comparisons between the 2005 and 2007 mappings in reading and mathematics at grades 4 and 8 were made separately for states that made changes in their testing systems and for those that made no such changes.[3] Key findings of the analysis presented in Section 4 are:

- In grade 4 reading, 12 of the 34 states with available data in both years indicated substantive changes in their assessments. Of those, eight showed significant differences between the 2005 and 2007 estimates of the NAEP scale equivalent of their state standards, half of which showed an increase and half a decrease.
- In grade 8 reading, 14 of the 38 states with available data in both years indicated substantive changes in their assessments. Of those, seven showed significant differences between the 2005 and 2007 estimates of the NAEP scale equivalent of their state standards, all seven showed lower 2007 estimates of the NAEP scale equivalents.

[3] The 2005 mappings in this report will not necessarily match previously published results (U.S. Department of Education 2007). Methodological differences between the procedures used in both analyses will generally cause empirical results to show small differences that are not large enough to change the whole-number scale value reported as the NAEP equivalent.

- In grade 4 mathematics, 14 of the 35 states with available data in both years indicated substantive changes in their assessments. Of those, 11 showed significant differences between the 2005 and 2007 estimates of the NAEP scale equivalent of their state standards: 6 states showed a decrease and 5 showed an increase.
- In grade 8 mathematics, 18 of the 39 states with available data in both years indicated substantive changes in their assessments. Of those, 12 showed significant differences between the 2005 and 2007 estimates of the NAEP scale equivalent of their state standards: 9 showed a decrease and 3 showed an increase.

For the states with no substantive changes in their state assessments in the same period, the analyses presented in Section 4 indicate that for the majority of states in the comparison sample (14 of 22 in grade 4 reading, 13 of 24 in grade 8 reading, 15 of 21 in grade 4 mathematics and 14 of 21 in grade 8 mathematics), the differences in the estimates of NAEP scale equivalents of their state standards were not statistically significant.

To address the third question, NAEP and state changes in achievement from 2005 to 2007 were compared. The percentage of students reported to be meeting the state standard in 2007 is compared with the percentage of the NAEP students in 2007 that is above the NAEP scale equivalent of the same state standard in 2005. The analysis was limited to states with (a) available data in both years and (b) no substantive changes in their state tests. The number of states included in the analyses ranged from 21 to 24, depending on the subject and grade. The expectation was 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 2005 and 2007. A more positive change on the state test indicates 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 students gained more on the NAEP skill domain. For example, 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. Key findings from Section 5 are:[4]

- In grade 4 reading, 11 of 22 states showed no statistically significant difference between NAEP and state assessment measures of changes in achievement; 5 states showed changes that are more positive than the changes measured by NAEP, and 6 states showed changes that are less positive than those measured by NAEP.
- In grade 8 reading, 9 of 24 states showed no statistically significant difference between NAEP and state assessment measures of achievement changes; 10 states showed changes that are more positive than the changes measured by NAEP, and 5 states showed changes that are less positive than those measured by NAEP.
- In grade 4 mathematics, 13 of 21 states showed no statistically significant difference between NAEP and state assessment measures of achievement changes; 5 states showed changes that

[4] Because differences between changes in achievement measured by NAEP and changes measured by the state assessment and the NAEP scale equivalents are based on the same data but are analyzed in different ways, statistically significant differences can be found in one and not the other because of the nonlinear relationship between scale scores and percentiles.

are more positive than the changes measured by NAEP, and 3 states showed changes that are less positive than those measured by NAEP.

- In grade 8 mathematics, 9 of 21 states showed no statistically significant difference between NAEP and state assessment measures of achievement changes, 7 states showed changes that are more positive than the changes measured by NAEP, and 5 states showed changes that are less positive than those measured by NAEP.

In considering the results described above, the reader should note that state assessments and NAEP are designed for different, though related purposes. 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 for summary assessment at an aggregate level. NAEP's achievement levels are used to interpret the meaning of the NAEP scales. 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.

In conclusion, these 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 states whether the rigor of their proficiency standards as represented by NAEP scale equivalents changed from 2005 to 2007. Significant differences in NAEP scale equivalents might reflect changes in state assessments and standards and/or other changes such as changes in policies or practices that occurred between the years. Finally, when key aspects of a state's assessment or standards remained the same, these mapping analyses allow NAEP to corroborate state-reported changes in student achievement and provide states with an indicator of the construct validity and generalizability of their test results.

CONTENTS

Page

Foreword iii

Executive Summary v

List of Tables xi

List of Figures xv

1 Introduction 1

Using NAEP to compare state performance standards 1

Data sources 3

Organization of this report 3

2 Estimation Methods 5

Estimation of the placement of state performance standards on the NAEP scale 6

Relative error 7

Measurement error in comparing NAEP and state measures of change 10

3 Mapping State Performance Standards 14

Reading 14

Mathematics 18

Cross-state comparisons 21

4 Comparing 2007 With 2005 State Performance Standards 26

Reading 27

Mathematics 34

5 Corroborating State Assessment Measures of Achievement Change With NAEP 41

6 Conclusions 47

7 References 49

Appendix A
Number of Schools in the NAEP Sample and the Percentage of Schools Used in the 2007 Mapping A-1

Appendix B
Changes in States' Assessments Between 2005 and 2007 ..... B-1

Appendix C
Supplementary Tables C-1

LIST OF TABLES

Table Page

1. Estimated NAEP scale equivalent scores for the state grades 4 and 8 reading proficient standards, their standard error and relative error, by state: 2007 16
2. Estimated NAEP scale equivalents scores for the state grades 4 and 8 mathematics proficient standards, their standard error and relative error, by state: 2007 19
3. Frequency of correlations between NAEP and state assessment school-level percentages meeting the proficient standards for reading and mathematics, grades 4 and 8: 2007 21
4. Correlations between NAEP and state assessment school-level percentages meeting the proficient standard for reading and mathematics grades 4 and 8, by state: 2007 22
5. Relationship between the percentage of students scoring proficient on the state test and the difficulty of grades 4 and 8 state standards as measured by the state's respective NAEP scale equivalent, by subject: 2007 23
6. Relationship between the percentage of students scoring proficient on NAEP and the difficulty of grades 4 and 8 state standards as measured by the state's respective NAEP scale equivalent t, by subject: 2007 25
7. State assessment data availability and state reports of whether 2005 and 2007 assessment results are comparable in grades 4 and 8 reading, by state: 2005 and 2007 28
8. States with both 2005 and 2007 data suitable to implement the mapping of grades 4 and 8 state reading standards, by whether the reported results are directly comparable 29
9. Number of NAEP schools, percentage of NAEP schools available for comparing state assessment results with NAEP results in grade 4 reading, and percentage of the student population represented in these comparison schools, by state: 2005 and 2007 30
10. Number of NAEP schools, percentage of NAEP schools available for comparing state assessment results with NAEP results in grade 8 reading, and percentage of the student population represented in these comparison schools, by state: 2005 and 2007 31
11. Difference between the estimated NAEP scale equivalents of state grade 4 reading proficient standards and their standard error, by state: 2005 and 2007 32
12. Difference between the estimated NAEP scale equivalents of state grade 8 reading proficient standards and their standard error, by state: 2005 and 2007 33
13. State assessment data availability and state reports of whether 2005 and 2007 assessment results are comparable in grades 4 and 8 mathematics, by state: 2005 and 2007 35
14. States with both 2005 and 2007 data suitable to implement the mapping of grades 4 and 8 mathematics standards, by whether the reported results are directly comparable 36
15. Number of NAEP schools, percentage of NAEP schools available for comparing state assessment results with NAEP results in grade 4 mathematics, and percentage of the student population in these comparison schools, by state: 2005 and 2007 37
16. Number of NAEP schools, percentage of NAEP schools available for comparing state assessment results with NAEP results in grade 8 mathematics, and percentage of the student population in the comparison schools, by state: 2005 and 2007 38

Table Page

17. Difference between the estimated NAEP scale equivalents of state grade 4 mathematics proficient standards and their standard error, by state: 2005 and 2007..... 39

18. Difference between the estimated NAEP scale equivalents of state grade 8 mathematics proficient standards and their standard error, by state: 2005 and 2007..... 40

19. NAEP and state assessment percentages meeting the state grade 4 reading proficient standard in 2007 based on 2005 standards, by state 42

20. NAEP and state assessment percentages meeting the state grade 8 reading proficient standard in 2007 based on 2005 standards, by state 43

21. NAEP and state assessment percentages meeting the state grade 4 mathematics proficient standard in 2007 based on 2005 standards, by state..... 44

22. NAEP and state assessment percentages meeting the state grade 8 mathematics proficient standard in 2007 based on 2005 standards, by state..... 45

23. States showing changes in student achievement from 2005 to 2007 in their own tests that are corroborated by NAEP results in the same period, by subject and grade 46

24. States showing changes in student achievement from 2005 to 2007 in their own tests that are statistically significantly more positive than NAEP's, by subject and grade..... 46

25. States showing changes in student achievement from 2005 to 2007 in their own tests that are statistically significantly less positive than NAEP's, by subject and grade..... 46

A-1 Number of NAEP schools, percentage of NAEP schools available for comparing state assessment results with NAEP results in grades 4 and 8 reading, and the percentage of the student population in these comparison schools, by state: 2007..... A-2

A-2 Number of NAEP schools, percentage of NAEP schools available for comparing state assessment results with NAEP results in grades 4 and 8 mathematics, and percentage of the student population in these comparison schools, by state: 2007..... A-3

B-1 Selected changes to state reading assessments between the 2004–05 and the 2006–07 administrations, by state B-2

B-2 Selected changes to state mathematics assessments between the 2004–05 and the 2006–07 administrations, by state B-4

B-3 Comparability of the 2007 state assessment results in reading and mathematics at grades 4 and 8 with the 2005 reported results, by state..... B-6

C-1 NAEP and state assessment percentages meeting the state grade 4 reading proficient standards in 2007 based on 2005 standards, by state..... C-2

C-2 NAEP and state assessment percentages meeting the state grade 8 reading proficient standards in 2007 based on 2005 standards, by state..... C-3

C-3 NAEP and state assessment percentages meeting the state grade 4 mathematics proficient standards in 2007 based on 2005 standards, by state C-4

C-4 NAEP and state assessment percentages meeting the state grade 8 mathematics proficient standards in 2007 based on 2005 standards, by state C-5

C-5 Number of states according to the comparability of state-reported results between 2005 and 2007, by the statistical significance of the discrepancy between NAEP and state measures of gains in grades 4 and 8 reading C-6

Table Page

C-6 Number of states according to the comparability of state-reported results between 2005 and 2007, by the statistical significance of the discrepancy between NAEP and state measures of gains in grades 4 and 8 mathematics C-6

C-7 Selected changes to state reading assessments between 2005 and 2007, by whether reports of grade 4 reading achievement changes from 2005 to 2007 in the state test and NAEP agree, by state C-7

C-8 Selected changes to state reading assessments between 2005 and 2007, by whether reports of grade 8 reading achievement changes from 2005 to 2007 in the state test and NAEP agree, by state C-8

C-9 Selected changes to state mathematics assessments between 2005 and 2007, by whether reports of grade 4 mathematics achievement changes from 2005 to 2007 in the state test and NAEP agree, by state C-9

C-10 Selected changes to state mathematics assessments between 2005 and 2007, by whether reports of grade 8 mathematics achievement changes from 2005 to 2007 in the state test and NAEP agree, by state C-10

LIST OF FIGURES

Figure Page

1. Mapping state proficiency standards onto the NAEP scale 6

2. NAEP scale equivalent scores for the state grades 4 and 8 reading standards for proficient performance, by state: 2007 17

3. NAEP scale equivalent scores for the state grades 4 and 8 mathematics standards for proficient performance, by state: 2007 20

4. Relationship between the percentage of students scoring proficient on NAEP and the difficulty of grades 4 and 8 state standards for reading and mathematics as measured by the state's respective NAEP scale equivalent: 2007 24

1 INTRODUCTION

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 been sponsoring research that focuses on comparing the proficiency standards of NAEP and states.[1] By mapping each state's standard for *proficient* performance onto the NAEP achievement scale, state policymakers can make comparisons of standards across states, in terms of the level of achievement required for *proficient* performance.

Recent studies that map state performance standards onto the NAEP scale have underlined the need for ongoing scrutiny of comparisons between NAEP and state assessment results.[2] In this report, we examine the consistency of the mapping results by using data from the state assessments and NAEP in 2005 and 2007. We investigate the impact and implications of the outcomes of the mapping procedure by using multiple years of data.

At a time when states are working to ensure that all their students reach proficient levels of achievement by 2014, as required by the No Child Left Behind Act of 2001 (NCLB), the analyses described in this report allow state policymakers to assess how high their state has set the bar for proficiency.

The comparison of achievement presented in this report is not intended to suggest deficiencies either in state assessments or in NAEP. The NAEP scales in reading and mathematics are being used as a common metric, not as a standard for evaluating state scales. Similarly, the NAEP achievement levels are provided simply as a national reference point for comparisons, not as a replacement for any given state's duly adopted state standards. Moreover, as provided by law, NCES, upon review of congressionally mandated evaluations of NAEP, has determined that NAEP achievement levels are to be used on a trial basis and should be interpreted with caution.[3] State-NAEP comparisons can help in the interpretation of state assessment results by providing a benchmark by which to assess changes in achievement that are measured by state assessments.

Using NAEP to compare state performance standards

The percentage of students identified as proficient on state assessments varies across states. Because each state's standard for proficient performance is set independently, the standards in different states can be quite different, even though they use the same terminology. A student who scores proficient in one state can move to another state and find that his or her performance is below the proficient range in the new state. NAEP, however, can provide the needed link to compare these assessment results across states. This comparison places all states' reading and mathematics standards on a common scale—the NAEP reading or mathematics scale—along

---

[1] Reports on this research are available at http://nces.ed.gov/nationsreportcard/studies/statemapping.asp.

[2] In early investigations, McLaughlin and Bandeira de Mello (2002, 2003, 2006) and subsequent reports (McLaughlin et al. 2008a, 2008b) mapped state primary performance standards onto the NAEP scale. Braun and Qian (2007) used similar methodology and data to conduct similar mappings. The recent mapping report from the National Center for Education Statistics is an outgrowth of these studies (U.S. Department of Education 2007).

[3] The status of NAEP achievement levels is available at http://nces.ed.gov/nationsreportcard/achlevdev.asp?.

with the NAEP achievement-level cut points. In this way, stakeholders can compare the relative stringency of state standards for proficiency in reading and mathematics.

A number of studies present arguments against the appropriateness of state assessment and NAEP comparisons. Criticisms of previous comparisons of state assessments and NAEP are not without merit and deserve thoughtful consideration. Prior criticisms of mapping studies have focused on three main topics: (1) state assessments and NAEP are developed for different purposes and have different goals and, as a result, should not be placed on a common scale; (2) state assessments may measure different constructs (e.g., Language Arts vs. Reading vs. Word Recognition) and should not be compared with one another; and (3) mapping studies implicitly use NAEP as the standard against which state assessments are ultimately determined to be deficient (Ho and Haertel 2007). Two National Research Council–sponsored studies have concluded that for a variety of reasons, mappings at the student level cannot be constructed validly (Feuer et al. 1999; Koretz, Bertenthal, and Green 1999).

Importantly, these studies do not address the appropriateness of mapping at the school level for the purpose of analyzing state-level results, which is the aim of the study described here. In a recent critique, Ho and Haertel (2007) posit that "substantial differences between state tests and NAEP will render the mapping illogical and subject to drift over time" (p. 1). If the standard that students must meet is not the same in 2007 as it was in 2005 (i.e., the standard "drifted"), then we cannot know whether achievement was better in one year or the other just because the percent achieving the standards was higher or lower in one year than in the other. Therefore, it is important to analyze NAEP and state assessment changes in achievement from 2005 to 2007, as was done for this report, which can determine whether there was drift. Drift indicates changes in either the state test or in NAEP (or in both) between consecutive administrations.

In an early mapping study, McLaughlin and Bandeira de Mello (2002, 2006) list a number of important caveats intended to prevent the misinterpretation of mapping results. They state emphatically that their report, among other things, (a) does not address questions about content, format, or administration of state assessments, as compared to NAEP, and (b) is not an evaluation of state assessments. As pointed out above, state assessments and NAEP are designed for different, although overlapping, purposes. For example, in many cases, state assessments are designed to provide pedagogical information about individual students to their parents and teachers, whereas NAEP is designed for summary assessment at an aggregate level. Findings of different standards, different trends, and different gaps should be presented without any implication that they be considered deficiencies either in state tests or in NAEP. However, it would be premature to conclude that two tests measuring grade 4 reading proficiency would assess no overlapping skills. Two tests that look quite different can measure the same variation because the various parts of reading (or mathematics) ability are highly correlated with one another. The high and consistent school-level correlations between state and NAEP assessment results suggests that state assessments and NAEP measure similar or related skills (McLaughlin *et al.* 2008, 2008b).

Despite the criticisms of NAEP and state assessment comparisons, there is a need for reliable information that compares state standards. What does it mean to say that a student is proficient in reading in grade 4 in Massachusetts? Would a fourth-grader who is proficient in reading in Wyoming be proficient in Oklahoma? However difficult it may be to answer these questions

definitively, they are fair questions that deserve consideration. In this study we examine the consistency of state standards when mapped onto the NAEP scale from 2005 to 2007.

Data sources

The analyses in the report are based on NAEP and state assessment results of public schools that participated in NAEP, weighted to represent the states.[4] The analyses use data from these sources: (a) NAEP data files for the states participating in the 2005 and 2007 reading and mathematics assessments, (b) state assessment school-level files compiled in the National Longitudinal School-Level State Assessment Score Database (NLSLSASD), and (c) school-level achievement data for the 2006-07 school year from ED*Facts*.[5] This report also relies on a review of state assessment programs conducted to gain contextual information about the general characteristics of state assessment programs and to help identify changes in states' assessments between the 2004-05 and 2006-07 school years that could affect the interpretation of the mapping results.[6]

The analyses presented are based on the standard NAEP estimates, which do not represent the achievement of those students with disabilities and/or English language learners who are excluded from NAEP testing.

Organization of this report

The report presents mapping results using the 2005 and 2007 NAEP assessments in mathematics and reading for grades 4 and 8. The analyses conducted for this study addresses the following questions:

- How do states' 2007 standards for proficient performance compare with each other when mapped onto the NAEP scale?
- How do the cut points on the NAEP scale that are equivalent to the scores required to meet a state's standard in 2007 compare to those estimated for 2005?
- Using the 2005 NAEP scale equivalent standards to define a state's *proficient* level of performance on NAEP, do NAEP and that state assessment agree on the changes in the proportion of students meeting that state's standard for proficiency from 2005 to 2007?

Section 2 of this report provides a description of the estimation methods used in the mapping and in the comparisons of results between 2005 and 2007. Section 3 presents the results of the analyses that examined the mapping results for 2007 in reading and mathematics at grades 4 and 8. Addressing the second question, Section 4 focuses on the comparison between the 2005 and 2007 mappings in reading and mathematics at grades 4 and 8. Addressing the third question,

---

[4] The method for sampling private schools in NAEP precludes using private school results in state-related reports. All NAEP published statistics at the state level are therefore for public schools only. Also, because private schools are not required to participate in a state's annual academic assessments under NCLB, private school data are not generally included in state test score databases.

[5] ED*Facts* is a collaborative effort among the U.S. Department of Education, State Education Agencies, and industry partners to centralize state–reported data into one federally coordinated, K–12 education data repository, located in the U.S. Department of Education.

[6] State profiles based on the *2007 Survey of State Assessment Program Characteristics* are available at http://nces.ed.gov/nationsreportcard/studies/statemapping.asp.

Section 5 discusses the NAEP and state assessment changes in achievement from 2005 to 2007, including possible explanations for discrepancies in the gains measured by the state tests and NAEP so that attention can be turned to identifying the sources of those discrepancies. Tables in appendix A show the sample sizes and percentages of the 2007 NAEP samples used in the analyses. Tables in appendix B summarize selected changes in states' assessments between the two NAEP administrations of 2005 and 2007 that could affect the interpretation of the mapping results. Appendix C includes tables with results complementing those discussed in the body of the report.

2 ESTIMATION METHODS

State assessment scores are usually reported as percentages of students in a grade at a school whose test scores are sufficiently high to meet a predefined state standard. That standard has been shown to vary a great deal from state to state (McLaughlin and Bandeira de Mello 2003). As a result, comparisons of percentages of students meeting state standards in different states are as much, if not more, a function of the placement of the standards as they are of differences in the achievement of the students.

Of essence in any attempt to compare changes in achievement on two tests is an understanding that the increase in the percentage of students meeting the standard depends critically on the placement of the standard. Generally, standards placed near the median test score (or more specifically, the modal test score) show the most increase in percentages meeting the standard, whereas relatively high and low standards lead to smaller changes in percentages meeting the standard (McLaughlin and Bandeira de Mello 2003). However, there are exceptions to this generality. For example, if instruction focuses on a particular subgroup of students located at one end of the distribution, a standard set at that end may show larger changes than a standard set in the middle.

To account for variation in the placement of standards on a scoring scale, the first step in comparing NAEP and state assessment measures of change is to measure changes in NAEP performance the same way change is measured on state assessments, that is, using the percentage of students in the state meeting that state's standard.[7] The process is done by *mapping* each state's standard onto the NAEP scale; that is, finding the NAEP scale value for the NAEP sample in the state for which the estimated percentage of students with higher NAEP scale values matches the percentage of students reported by the state as achieving the state's standard in the same schools. Of course, because NAEP is based on a sample of students in each participating school, and because both assessments have measurement error, there is some mapping error in determining the NAEP equivalent of a state's standard. It is necessary to consider mapping error for valid comparisons between NAEP and state assessment measures of change in achievement.

This section summarizes the estimation methods used in the mapping procedure to place state performance standards onto the NAEP scales and in the comparison analysis between 2005 and 2007. We develop a framework for evaluating differences between achievement changes measured by NAEP and by state tests. Essentially, NAEP and state achievement changes in each subject and grade are rendered comparable by summarizing NAEP results in a state as the percentage meeting the state's standard, which requires, as a first step, mapping the state's standard onto the NAEP scale.

[7] Given that the only test results systematically available for all states are percentages of students in each school with scores higher than a cut point (i.e., *meeting the standard*), finding the NAEP equivalent of that cut point is an essential step in comparing achievement gains based on state test data to achievement gains on NAEP. If state test means and standard deviations were available for schools in the NAEP sample, mapping of the standards, while important in itself, would not be required for comparing state test and NAEP achievement gains.

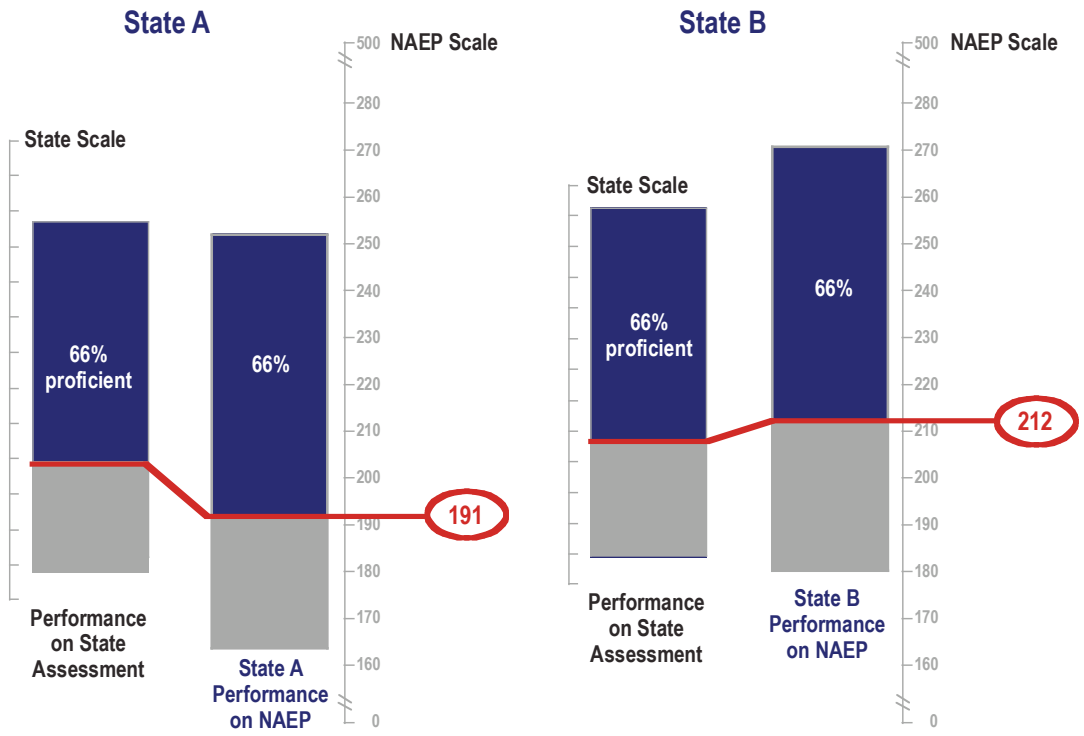
Estimation of the placement of state performance standards on the NAEP scale

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

1. 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.
2. 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.
3. 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.
4. 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).

Using figure 1 to illustrate, we see that 66 percent of the students in State A meet that state's standard (estimated from step 2); based on State A's NAEP sample, 66 percent of State A's students score above 191 on the NAEP scale (using the distribution obtained in step 3). Suppose that in State B, where students perform higher on NAEP than in State A, 66 percent of its students also meet its state standard. This translates into a higher NAEP scale equivalent (212 in the illustration), because 66 percent of State B's students score above 212 on the NAEP scale, based on State B's NAEP sample. State A's standard corresponds to, or maps onto, a lower level of NAEP achievement than State B's standard does, even though each state reports the same 66 percent meeting its own standard.

Figure 1. Mapping state proficiency standards onto the NAEP scale



The reported percentage meeting the state's standard in each NAEP school s, p_s (e.g., 66 percent as in figure 1), 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.[8] 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:[9]

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

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

where the sum is over students in schools participating in NAEP, and $\partial_{isv}(c)$ is an indicator variable that is 1 if the vth 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.[10] 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.[11,12]

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.

---

[8] 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.

[9] 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.

[10] Appendix A of McLaughlin et al. (2008a) describes in more detail the technical aspects of the placement of state achievement standards on the NAEP scale.

[11] NAEP computes standard error using a combination of *sampling error* based on Jackknife resampling and *measurement error* from the variance between plausible values.

[12] This mapping procedure is analogous to the one used in U.S. Department of Education (2007) and produces results that are qualitatively similar. The distinctions between the two procedures are discussed in Braun and Qian (2007).

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 $K < k$,

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

where $0 \leq k \leq 1$.

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

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

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.

Measurement error in comparing NAEP and state measures of change

Under No Child Left Behind, each state has developed measurements for determining whether its schools are making adequate yearly progress (AYP), which refers not to the progress of a child from, say, fourth grade to fifth grade but to the progress of a school in increasing the performance of its fourth-graders from one year to the next. The basic idea of comparing achievement changes from one year's students in a particular grade with achievement changes from another year's students in the same grade is that a set of skills is to be learned and that these skills might be more (or less) thoroughly learned by the students in one year than they were by the students in the other year. A test is written that samples the skill domain and is given to each of the two cohorts of students, and the scores are compared. Of course, the average scores will not be exactly the same in the two years if the test merely samples the skill domain and does so on a finite number of students. However, a simple statistical test can be executed to determine whether the difference is in the realm of random variation. If the sample of students were infinitely large and the test measured all the skills in the domain without error, the standard errors would be *zero*, meaning that any difference between the scores of the two cohorts would be statistically significant. Whether a difference is *important* is another question, but differences that are not statistically significant should not be considered further because they may well reflect just chance variation. [13]

Letting D be the discrepancy between changes from year 1 to year 2 in percentage meeting the state standard identified by the state test and the changes in the same period in the same percentages when measured by NAEP, we can test for whether D is statistically significantly different from zero by estimating the ratio of D to its standard error. However, to interpret the results of such a comparison, we also need to consider the explanations of statistically significant values of D. These discrepancies represent an additional source of error that contributes to the differences in achievement changes identified by NAEP and by the state assessment program. In general, such differences are hypothesized to be the result of some systematic difference between what the state assessment measures and what NAEP measures (in test content, student populations, or test administration). We call this a *true score error* to distinguish it from discrepancies arising from the finiteness of the samples and the imperfections of measurement.[14]

[13] The following discussion is excerpted from a report to NCES on the measurement error in comparing NAEP and state test gains (McLaughlin 2008).

[14] One source of error is due to the systematic differences in the domains of skills assessed by NAEP and the state assessment, and not to random measurement error or to sampling error. A second kind of error arises because both tests measure the domain with some error and because the mapping is based on a finite sample of students. The distribution of NAEP scores in the sample of NAEP students in the specified schools is likely to be slightly different from the hypothetical distribution of NAEP achievement of all students tested by the state in those schools, leading to small over- or underestimates of the NAEP scale equivalent of the state standard.

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

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

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 $D = (60 - 55) - (50 - 50) = 5$. 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 $(\hat{p}_{2S} - \hat{p}_{2N|map=1})$ 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.[15] 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

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

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

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.[16] 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.

Measuring the standard error of the mapping

Estimating the standard error of the mapping is not a necessary step in determining the standard error of D because we can apply the NAEP jackknife technique directly to the estimate of D. However, 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,

$$\hat{c} = \hat{c}_1 - \hat{c}_2, \qquad [7]$$

is just the square root of the sum of the squares of the standard errors of the two separate NAEP scale equivalents. That is,

$$SE(\hat{c}) = \sqrt{SE(\hat{c}_1)^2 + SE(\hat{c}_2)^2}\,. \qquad [8]$$

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

Summary

The ultimate purpose for estimating the standard error of D is to decide whether differences between changes in achievement showed by NAEP and changes in achievement showed by the state are sufficiently large that they are not likely to be due to random factors. If the difference, D, is statistically significantly different from zero, students gained more on either the NAEP skill domain or on the state-specific skill domain than represented by those domains' contributions to variance in year 1. Focusing on state-specific content during instruction might be expected to lead to a positive value for D, whereas focusing entirely on NAEP content might be expected to lead to a negative value for D. Other explanations for a larger change on the state test exist. A statistically significant value of D may be due to a change in the content, administration, or scoring of either the state test or NAEP in the interval. For example, a change in the NAEP exclusion rates between years (for whatever reason) can lead to a significant D; a larger apparent state gain (i.e., a positive change) could be due to increased familiarity with and focus on the state test in the schools, with teaching students how to do particular kinds of items on the state test; and decreasing the focus on some aspects of NAEP content in the state curriculum between the two assessments could lead to a larger gain on the state test.

[16] These factors are discussed in McLaughlin (2008).

The key underlying assumption is that NAEP and the state assessment each remain essentially the same over the two years. If either test is substantively changed between the two years, then comparisons of changes identified on the two tests are not warranted. NAEP did not go through any substantive methodological changes between 2005 and 2007. However, in the years from 2005 to 2007, the focus of this report, many states changed their state assessments to ensure that they were complying with the regulations of the NCLB law, and finding values of *D* significantly different from zero in those cases is to be expected.[17]

It should be noted that the state assessment data available for this study include only a single number (percentages) reported for each school (for each subject and grade). *D* is based on a match of NAEP and a state's assessment at a single point in the state's achievement distribution.

Finally, there is the question of what is meant by stability of the mapping between two years and how it can be measured. In practical terms, the value of *D* is a measure of the *instability* of the mapping. If $D = 0$, the mappings in the two years yield identical results. If *D* is positive (the state showed a more positive change than the change measured by NAEP), that means that if we were to calculate the NAEP scale equivalent of the standard in year 2, the result would be a lower value on the NAEP scale than the equivalent obtained from the year 1 mapping. This does not necessarily mean that the state's standard got easier. If both NAEP and the state's assessment and scoring systems remained constant over the 2-year interval, it means that there were more gains on state-specific skills than on NAEP skills during the interval.

[17] Tables in appendix B summarize selected changes in states' assessments between the two NAEP administrations of 2005 and 2007.

3 MAPPING STATE PERFORMANCE STANDARDS

The results of the mapping procedure are presented below for reading and mathematics. Three jurisdictions are not included in the 2007 analyses because data were unavailable: District of Columbia, Nebraska, and Utah. In addition, California grade 8 mathematics data were unavailable.

Sample sizes and percentages of the 2007 NAEP samples used in the analyses are shown in appendix A. In some states, the student population represented by NAEP is less than 100 percent of the total population because state assessment scores are missing for some schools. Scores may be missing because of either the failure to match schools in the NAEP and state databases or the suppression of scores where there are too few students. Overall, with the exception of Wisconsin in both subjects and grades, the estimated percentages of the student population represented by the schools used in the analyses are at least 90 percent.[18]

Reading

Table 1 displays the NAEP scale equivalents of each state's reading standards for *proficient* performance for grades 4 and 8. Standard errors of the NAEP scale equivalent estimates and the relative error criterion, *K*, a measure of how well the procedure reproduces the percentages reported by the state to be meeting the standard in each school in the NAEP sample, are also included. As previously discussed, the criterion proposed is to consider relative errors greater than .5 as indicating that the mapping error is too large to support useful inferences from the placement of the standard on the NAEP scale without any additional evidence. Only one grade 4 reading standard (Texas) and one grade 8 reading standard (Virginia) have relative errors greater than .5. The within-school discrepancies between NAEP and Indiana grade 8 test results seem to be smaller than the discrepancies that we would expect owing to NAEP student within-school sampling error alone.[19]

In 2007, states' standards for proficient performance in reading varied greatly in difficulty as reflected in their NAEP scale equivalent scores. The NAEP scale equivalents of states' proficient standards ranged from below the NAEP *Basic* level to the NAEP *Proficient* level (see figure 2).

In reading, at grade 4, the average of the estimated standards for proficiency across states was equivalent to a score of 199 (data not shown) on the NAEP scale, below the NAEP cut point for *Basic* performance (208). Taking the standard errors into account, the estimated difference between the five states with the highest standards and the five states with the lowest standards was at least 29 points on the NAEP scale, comparable to the 30-point distance between the NAEP *Basic* standard (208) and the NAEP *Proficient* standard (238). Another way of looking at it is that the distance separating the five most difficult standards to achieve and the five least difficult standards to achieve was under one standard deviation in student performance on the

[18] For Wisconsin, the grade 4 reading and mathematics analyses are based on 65 percent of the NAEP schools serving about 71 percent of the students represented by NAEP. Analyses for grade 8 reading and mathematics are based on 75 percent of the NAEP schools, serving about 83 percent of the students represented by NAEP.

[19] Because the relative error is actually a sample statistic with its own random variation and because it can take on negative values (if the differences between school means on NAEP and the state test are smaller than would be expected given within-school sample sizes), those negative values are displayed with the § symbol.

grade 4 NAEP (36 points). Accounting for the margin of error, 31 of the 48 states set grade 4 standards for proficiency (as measured on the NAEP scale) that were lower than the *Basic* performance on NAEP (208).

For grade 8 reading, the average NAEP scale equivalent score was 246 (data not shown), above the NAEP cut point for *Basic* performance (243). The variation among states at grade 8 was as large as the variation at grade 4. The estimated difference between the five states with the highest standards and the five states with the lowest standards was at least 29 points on the NAEP scale (also taking the standard error into account), less than the 38-point distance between *Basic* (243) and *Proficient* performance (281) on NAEP, and below the one standard deviation in student performance on the grade 8 NAEP (35 points). Accounting for the margin of error, 15 of the 48 states set grade 8 standards for proficiency (as measured on the NAEP scale) that were lower than the *Basic* performance on NAEP.

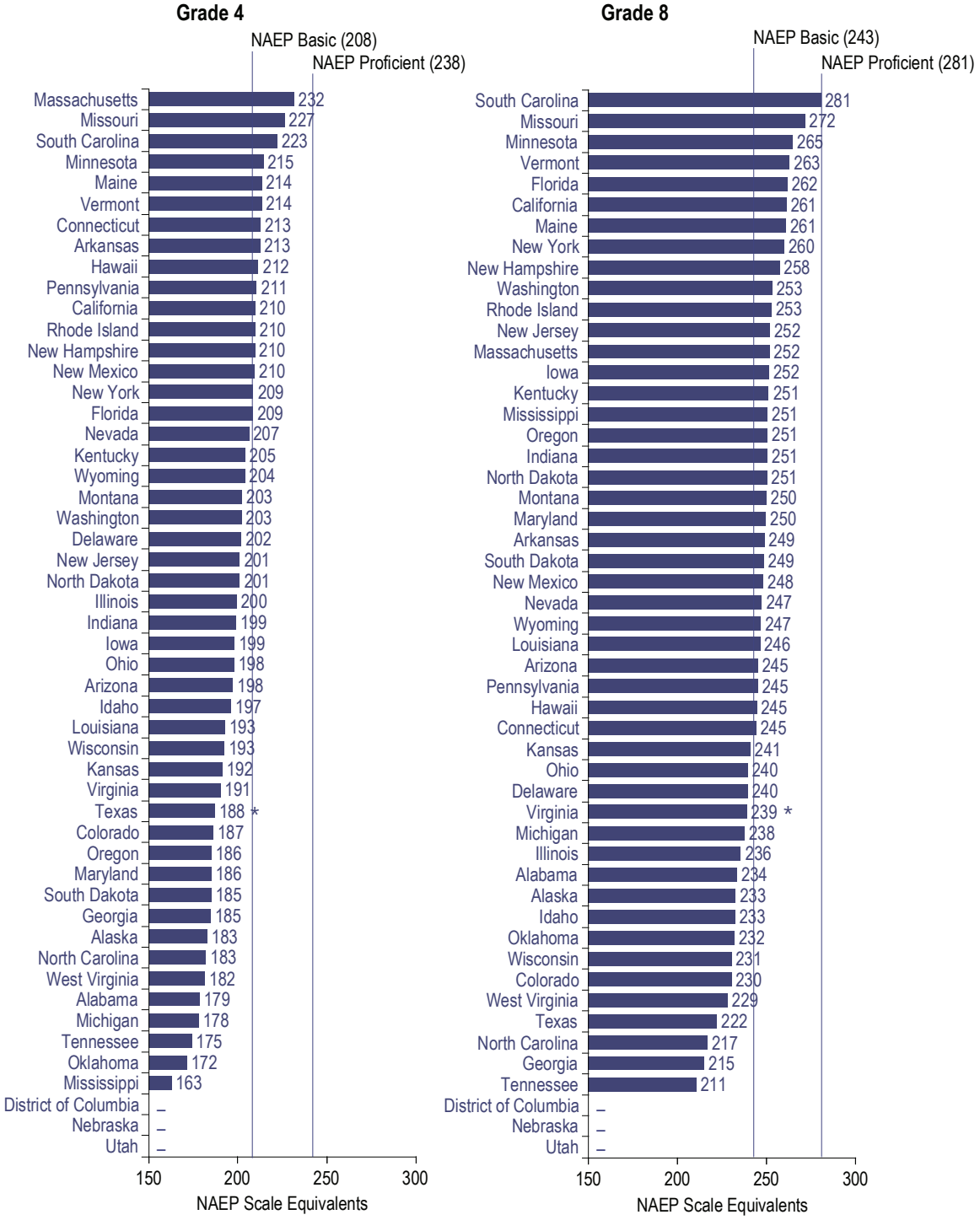
In reading, Missouri, Minnesota, and South Carolina were among the five states with the most difficult standards for proficiency at both grade levels. Tennessee appears among the five states with the least difficult standards at both grade levels.

Table 1. Estimated NAEP scale equivalent scores for the state grades 4 and 8 reading proficient standards, their standard error and relative error, by state: 2007

State/jurisdiction	Grade 4 NAEP scale equivalent	Grade 4 Standard error	Grade 4 Relative error[1]	Grade 8 NAEP scale equivalent	Grade 8 Standard error	Grade 8 Relative error[1]
Alabama	179	1.5	0.4	234	1.5	0.2
Alaska	183	0.9	0.1	233	1.9	0.2
Arizona	198	1.4	0.1	245	1.1	#
Arkansas	213	1.4	0.2	249	1.4	0.4
California	210	0.9	0.1	261	0.6	#
Colorado	187	1.5	0.1	230	1.4	0.1
Connecticut	213	1.6	0.1	245	1.1	#
Delaware	202	0.9	0.3	240	1.0	0.4
District of Columbia	—	†	†	—	†	†
Florida	209	0.8	0.1	262	0.8	#
Georgia	185	1.3	0.5	215	1.7	0.4
Hawaii	212	1.0	0.2	245	0.7	0.1
Idaho	197	1.4	0.4	233	1.0	#
Illinois	200	1.4	0.3	236	1.5	0.5
Indiana	199	1.3	0.1	251	0.7	§
Iowa	199	1.7	0.4	252	1.1	0.1
Kansas	192	1.9	0.3	241	1.0	0.3
Kentucky	205	1.6	0.3	251	1.1	0.3
Louisiana	193	2.2	0.5	246	1.3	0.2
Maine	214	1.0	0.2	261	0.9	0.3
Maryland	186	1.5	0.3	250	1.2	0.1
Massachusetts	232	1.2	0.2	252	1.1	0.1
Michigan	178	2.5	0.4	238	1.2	0.1
Minnesota	215	1.4	0.2	265	0.7	0.3
Mississippi	163	1.3	0.3	251	0.6	0.1
Missouri	227	1.1	0.3	272	1.1	#
Montana	203	1.2	0.4	250	1.5	0.3
Nebraska	—	†	†	—	†	†
Nevada	207	1.1	0.2	247	1.0	0.3
New Hampshire	210	0.8	0.4	258	1.5	0.4
New Jersey	201	2.0	0.2	252	1.1	0.1
New Mexico	210	0.7	0.3	248	1.0	0.1
New York	209	1.4	0.1	260	0.9	0.1
North Carolina	183	1.0	0.3	217	1.2	0.3
North Dakota	201	1.0	0.4	251	1.4	0.4
Ohio	198	2.2	0.4	240	1.9	0.2
Oklahoma	172	3.7	0.4	232	1.6	0.2
Oregon	186	2.1	0.4	251	1.2	0.3
Pennsylvania	211	1.2	0.1	245	1.4	0.1
Rhode Island	210	1.1	0.2	253	1.1	0.1
South Carolina	223	1.5	0.2	281	1.0	0.2
South Dakota	185	1.7	0.4	249	0.9	0.3
Tennessee	175	1.7	0.4	211	2.5	0.3
Texas	188	1.6	0.6	222	1.1	0.2
Utah	—	†	†	—	†	†
Vermont	214	1.0	0.5	263	1.4	0.4
Virginia	191	1.6	0.5	239	1.2	0.6
Washington	203	2.1	0.4	253	1.2	0.2
West Virginia[2]	182	1.4	0.3	229	1.3	0.4
Wisconsin[2]	193	2.0	0.3	231	1.4	0.2
Wyoming	204	1.2	0.5	247	1.1	0.5

— State assessment data not available.
† Not applicable.
# Rounds to zero.
§ The within-school discrepancies between NAEP and state test results are no larger, and possibly smaller, than discrepancies that would be expected owing to NAEP student within-school sampling error alone.
[1] Inferences based on estimates with relative error greater than .5 may require additional evidence.
[2] The percentage of the student population represented by the NAEP schools used in the estimations was less than 90 percent in at least one grade.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Figure 2. NAEP scale equivalent scores for the state grades 4 and 8 reading standards for proficient performance, by state: 2007



— State assessment data not available.

* Relative error greater than .5.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Mathematics

Table 2 displays the NAEP scale equivalent scores of each state's mathematics standards for proficient performance for grades 4 and 8. Standard errors of the NAEP scale equivalent estimates and the relative error criterion, *K*, are also included. Seven of the 48 grade 4 mathematics standards (Alabama, Georgia, Indiana, Michigan, New Hampshire, Oklahoma, and Virginia) have relative errors greater than .5 indicating that the variation in results for individual schools are large enough to call into question the use of these equivalents without additional supporting evidence. In grade 8, only Virginia has a mapping with relative error above .5. For two states, Connecticut and South Carolina, the within-school discrepancies between NAEP and state grade 8 test results are smaller than the discrepancies that we would expect owing to NAEP student within-school sampling error alone.

In mathematics at grade 4, the average NAEP scale equivalent across states was 223 (data not shown), about one-third of the way between the NAEP cut points for *Basic* (214) and *Proficient* (249) performance, as shown in figure 3. Taking the standard errors into account, the difference between the five states with the highest standards and the five states with the lowest standards was estimated to be 29 points on the NAEP scale, close to the distance between the NAEP *Basic* standard and the NAEP *Proficient* standard (35 points) and about a full standard deviation in grade 4 NAEP mathematics achievement (29 points). Accounting for the margin of error, 7 of the 48 states set grade 4 standards for proficiency (as measured on the NAEP scale) that were lower than the *Basic* performance level on NAEP, and 1 state set standards above the 249 NAEP *Proficient* cut point.

In mathematics at grade 8, the mean NAEP scale equivalent was 271 (data not shown) on the NAEP scale, above the NAEP cut point for *Basic* performance (262). The difference between the five states with the highest standards and the five states with the lowest standards was at least 29 points on the NAEP scale, less than the distance between the NAEP *Basic* standard and the NAEP *Proficient* standard (37 points) and close to one standard deviation in grade 8 NAEP mathematics achievement (36 points). Accounting for the margin of error, we see that 8 of the 47 states set grade 8 standards for proficiency (as measured on the NAEP scale) that were lower than the *Basic* performance on NAEP, and 2 states set standards above the 299 NAEP *Proficient* cut point.

In mathematics, Massachusetts, Missouri, South Carolina, and Washington were among the states with the most difficult standards at both grade levels in 2007. At both grade levels, Tennessee was the state with the least difficult standards.

Table 2. Estimated NAEP scale equivalent scores for the state grades 4 and 8 mathematics proficient standards, their standard error and relative error, by state: 2007

State/jurisdiction	Grade 4 NAEP scale equivalent	Grade 4 Standard error	Grade 4 Relative error[1]	Grade 8 NAEP scale equivalent	Grade 8 Standard error	Grade 8 Relative error[1]
Alabama	205	1.5	0.8	253	1.9	0.4
Alaska	216	1.3	0.3	265	1.2	0.3
Arizona	213	1.4	0.1	268	1.1	0.1
Arkansas	229	0.6	0.2	277	1.3	0.1
California	226	0.7	0.4	—	†	†
Colorado	201	1.6	0.2	259	1.3	0.1
Connecticut	220	0.7	0.1	252	2.0	§
Delaware	225	0.7	0.2	272	0.9	#
District of Columbia	—	†	†	—	†	†
Florida	230	0.8	0.2	266	0.9	#
Georgia	213	0.8	0.9	243	1.7	0.3
Hawaii	238	0.5	0.2	294	0.8	0.2
Idaho	217	0.9	0.5	265	1.6	#
Illinois	208	0.9	0.3	251	0.8	0.1
Indiana	228	0.9	0.6	266	1.6	0.1
Iowa	220	1.1	0.3	264	1.5	0.1
Kansas	219	1.3	0.5	270	1.6	0.4
Kentucky	229	1.0	0.4	279	0.7	0.2
Louisiana	223	1.3	0.3	267	1.2	0.1
Maine	236	0.8	0.2	286	0.9	0.1
Maryland	206	1.3	0.5	278	1.5	#
Massachusetts	254	1.0	0.3	302	1.1	0.1
Michigan	204	1.6	0.6	260	1.5	0.1
Minnesota	237	0.9	0.2	286	0.9	0.2
Mississippi	204	0.8	0.5	262	0.9	#
Missouri	245	0.8	0.4	289	1.2	0.1
Montana	234	1.0	0.3	281	1.7	0.1
Nebraska	—	†	†	—	†	†
Nevada	224	1.1	0.3	267	1.2	0.1
New Hampshire	239	1.1	0.6	282	0.8	0.3
New Jersey	220	1.1	0.4	272	0.8	0.1
New Mexico	233	0.8	0.3	285	0.9	0.1
New York	219	0.8	0.2	273	1.1	0.1
North Carolina	231	0.6	0.3	270	1.3	0.1
North Dakota[2]	226	1.0	0.4	279	0.8	0.3
Ohio	225	1.3	0.5	265	1.2	0.2
Oklahoma	213	1.5	0.8	249	1.1	0.3
Oregon	220	0.8	0.4	262	1.2	0.2
Pennsylvania	223	0.9	0.2	271	1.0	0.1
Rhode Island	236	0.7	0.1	279	0.6	#
South Carolina	245	0.9	0.2	312	1.4	§
South Dakota	224	1.0	0.2	271	0.7	0.1
Tennessee	198	1.3	0.4	234	2.2	0.4
Texas	217	0.9	0.5	268	1.0	0.2
Utah	—	†	†	—	†	†
Vermont[2]	239	1.0	0.3	284	0.9	0.1
Virginia	219	0.9	0.6	259	1.6	0.6
Washington	240	0.8	0.2	286	1.1	#
West Virginia[2]	217	1.3	0.4	253	1.0	0.1
Wisconsin[2]	222	2.3	0.2	262	1.7	0.1
Wyoming	216	0.6	0.5	279	0.8	0.4

— State assessment data not available.

† Not applicable.

Rounds to zero.

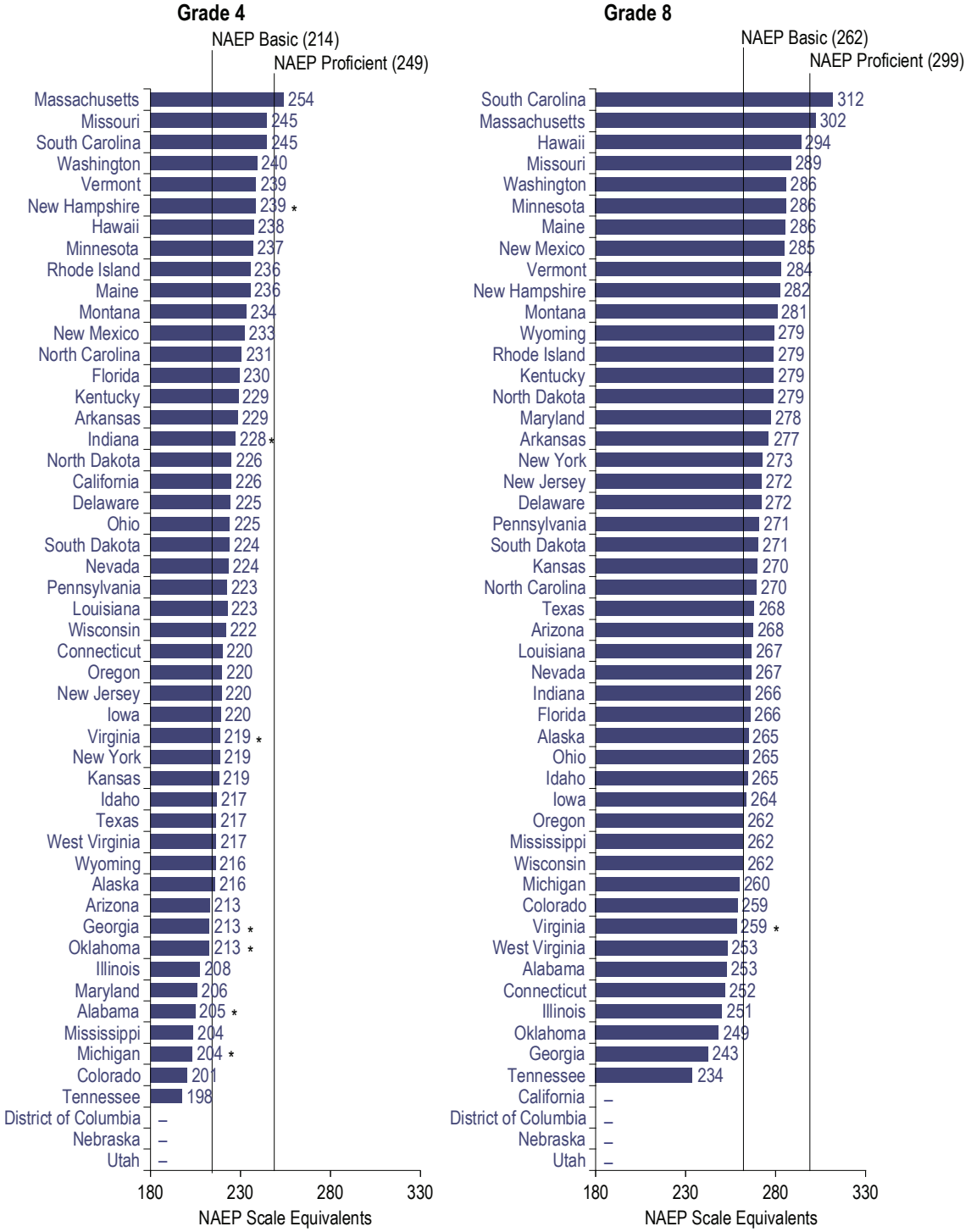
§ The within-school discrepancies between NAEP and state test results are no larger, and possibly smaller, than discrepancies that would be expected owing to NAEP student within-school sampling error alone.

[1] Inferences based on estimates with relative error greater than .5 may require additional evidence.

[2] The percentage of the student population represented by the NAEP schools used in the estimations was less than 90 percent in at least one grade.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Figure 3. NAEP scale equivalent scores for the state grades 4 and 8 mathematics standards for proficient performance, by state: 2007



— State assessment data not available.

* Relative error greater than .5.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Cross-state comparisons

The majority of the states included in the analyses had state assessment results that were correlated with NAEP, with correlations of .7 or more: that is, both assessments identified similar patterns of achievement across schools within the state.[20] The school-level correlations between the percentage of schools' students meeting the NAEP and the state assessment standards for proficiency are summarized in table 3 and listed by state in table 4.

Table 3. Frequency of correlations between NAEP and state assessment school-level percentages meeting the proficient standards for reading and mathematics, grades 4 and 8: 2007

	Reading		Mathematics	
Correlation	Grade 4	Grade 8	Grade 4	Grade 8
$.3 \leq r < .4$	0	1	0	0
$.4 \leq r < .5$	0	1	2	0
$.5 \leq r < .6$	7	9	3	3
$.6 \leq r < .7$	14	13	12	6
$.7 \leq r < .8$	13	11	18	22
$.8 \leq r < .9$	14	12	13	14
$.9 \leq r$	0	1	0	2
Number of states[1]	48	48	48	47

[1] Test data for the District of Columbia, Nebraska, and Utah were not available to be included in the analysis. California does not test grade 8 mathematics.

NOTE: Frequency counts are based on unrounded correlation coefficients as opposed to the rounded coefficients shown in table 4.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading and Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

In reading, at both grade levels, at least half the states had correlations of .7 or more. Correlations were higher in mathematics than in reading. In mathematics, 31 of the 48 states included in grade 4 and 38 of the 47 states in grade 8 had correlations of .7 or higher.

Although the majority of states reported assessment results that identified the same patterns of achievement across schools as did NAEP, a small number of states (ranging from 3 to 11 depending on subject and grade) had test results that did not correlate as well with NAEP results, with correlations of less than .6, as shown in tables 3. For example, from table 4, North Dakota, Oklahoma, West Virginia, and Wyoming had correlations below .6 on at least three of the four assessments. This could be the result of small enrollments in schools in these states which affect the reliability of the percentages of students meeting a standard. Another possible explanation is that the tests measure different things. It is possible that assessments that sample and measure different parts of the reading and mathematics domain might still be highly correlated; that is, they might still identify the same schools as high achieving and low achieving.[21,22] Nevertheless, the relatively low correlations in a few states need to be considered when we interpret the results of comparisons of NAEP and state assessment results.

[20] A correlation of .7 implies that 50% of the variance of one variable can be predicted from the other variable.

[21] A variety of factors can lead to low correlations between tests covering the same content: size of the school sample of students on which the percentage is based, conditions of testing, time of testing, motivation to perform, similarity of accommodations provided, match of the student populations included in the statistics, etc.

[22] Indiana, Iowa, Michigan, New Hampshire, North Dakota, Rhode Island, Vermont, and Wisconsin are states with testing in the fall and they may be measuring previous grade skills.

Table 4. Correlations between NAEP and state assessment school-level percentages meeting the proficient standard for reading and mathematics grades 4 and 8, by state: 2007

State/jurisdiction	Reading		Mathematics	
	Grade 4	Grade 8	Grade 4	Grade 8
Alabama	0.67	0.72	0.67	0.74
Alaska	0.81	0.81	0.75	0.78
Arizona	0.86	0.84	0.86	0.80
Arkansas	0.76	0.69	0.82	0.73
California	0.88	0.84	0.76	—
Colorado	0.84	0.75	0.80	0.80
Connecticut	0.90	0.90	0.90	0.90
Delaware	0.68	0.71	0.79	0.92
District of Columbia	—	—	—	—
Florida	0.80	0.81	0.81	0.82
Georgia	0.70	0.58	0.76	0.75
Hawaii	0.73	0.78	0.74	0.77
Idaho	0.59	0.68	0.61	0.74
Illinois	0.80	0.60	0.83	0.79
Indiana[1]	0.75	0.80	0.65	0.78
Iowa[1]	0.53	0.66	0.65	0.75
Kansas	0.60	0.65	0.60	0.61
Kentucky	0.67	0.63	0.65	0.72
Louisiana	0.71	0.71	0.79	0.83
Maine	0.64	0.54	0.75	0.72
Maryland	0.71	0.82	0.70	0.89
Massachusetts	0.80	0.82	0.75	0.86
Michigan[1]	0.71	0.79	0.78	0.88
Minnesota	0.73	0.65	0.78	0.72
Mississippi	0.65	0.80	0.67	0.81
Missouri	0.72	0.77	0.72	0.81
Montana	0.63	0.68	0.68	0.71
Nebraska	—	—	—	—
Nevada	0.82	0.70	0.82	0.78
New Hampshire[1]	0.61	0.60	0.63	0.69
New Jersey	0.82	0.84	0.77	0.87
New Mexico	0.74	0.71	0.75	0.79
New York	0.85	0.81	0.83	0.83
North Carolina	0.66	0.67	0.81	0.82
North Dakota[1]	0.63	0.50	0.59	0.58
Ohio	0.76	0.74	0.72	0.82
Oklahoma	0.59	0.56	0.43	0.53
Oregon	0.71	0.69	0.69	0.69
Pennsylvania	0.87	0.84	0.84	0.86
Rhode Island[1]	0.80	0.90	0.86	0.93
South Carolina	0.79	0.69	0.81	0.78
South Dakota	0.65	0.58	0.73	0.75
Tennessee	0.73	0.67	0.75	0.70
Texas	0.64	0.68	0.66	0.73
Utah	—	—	—	—
Vermont[1]	0.54	0.49	0.67	0.68
Virginia	0.56	0.55	0.60	0.63
Washington	0.68	0.68	0.85	0.79
West Virginia	0.56	0.38	0.59	0.55
Wisconsin[1]	0.82	0.81	0.87	0.85
Wyoming	0.56	0.53	0.45	0.65

— State assessment data not available.

[1] State with fall testing.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading and Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

In 2007, as was the case for the 2003 and 2005 mapping results, most of the variation between states in the proportion of proficient students on state assessments can be explained by the rigor of a state's standard for *proficient* performance. Table 5 shows the estimated linear relationships between the difficulty of each state's standard for proficiency, as measured by its NAEP scale equivalent and the percentage of students scoring *proficient* on the state test: states with a more difficult standard for proficiency (as measured on the NAEP scale) tend to have *fewer* students scoring proficient, whereas states with less difficult standards tend to have more students scoring proficient. The negative slopes of the lines fitted to the data points (states) show that each 1-point increase in the difficulty of a state's standard for proficiency in reading as measured by the NAEP scale is associated with .7 to .8 percentage point fewer students meeting the standards in grades 4 and 8, respectively. In mathematics, the relationship is similar.

Table 5. Relationship between the percentage of students scoring proficient on the state test and the difficulty of grades 4 and 8 state standards as measured by the state's respective NAEP scale equivalent, by subject: 2007

Percent proficient on state test = f(state standards as measured by the state's NAEP scale equivalent)							
		Grade 4			**Grade 8**		
Subject		**Intercept**	**Slope**	**R^2**	**Intercept**	**Slope**	**R^2**
Reading	Estimate	214.1	-.7 *	.70	272.7	-.8 *	.69
	Standard error	*13.49*	*.07*	†	*19.96*	*.08*	†
Mathematics	Estimate	268.5	-.9 *	.70	288.7	-.8 *	.71
	Standard error	*19.10*	*.09*	†	*21.68*	*.08*	†

† Not applicable.

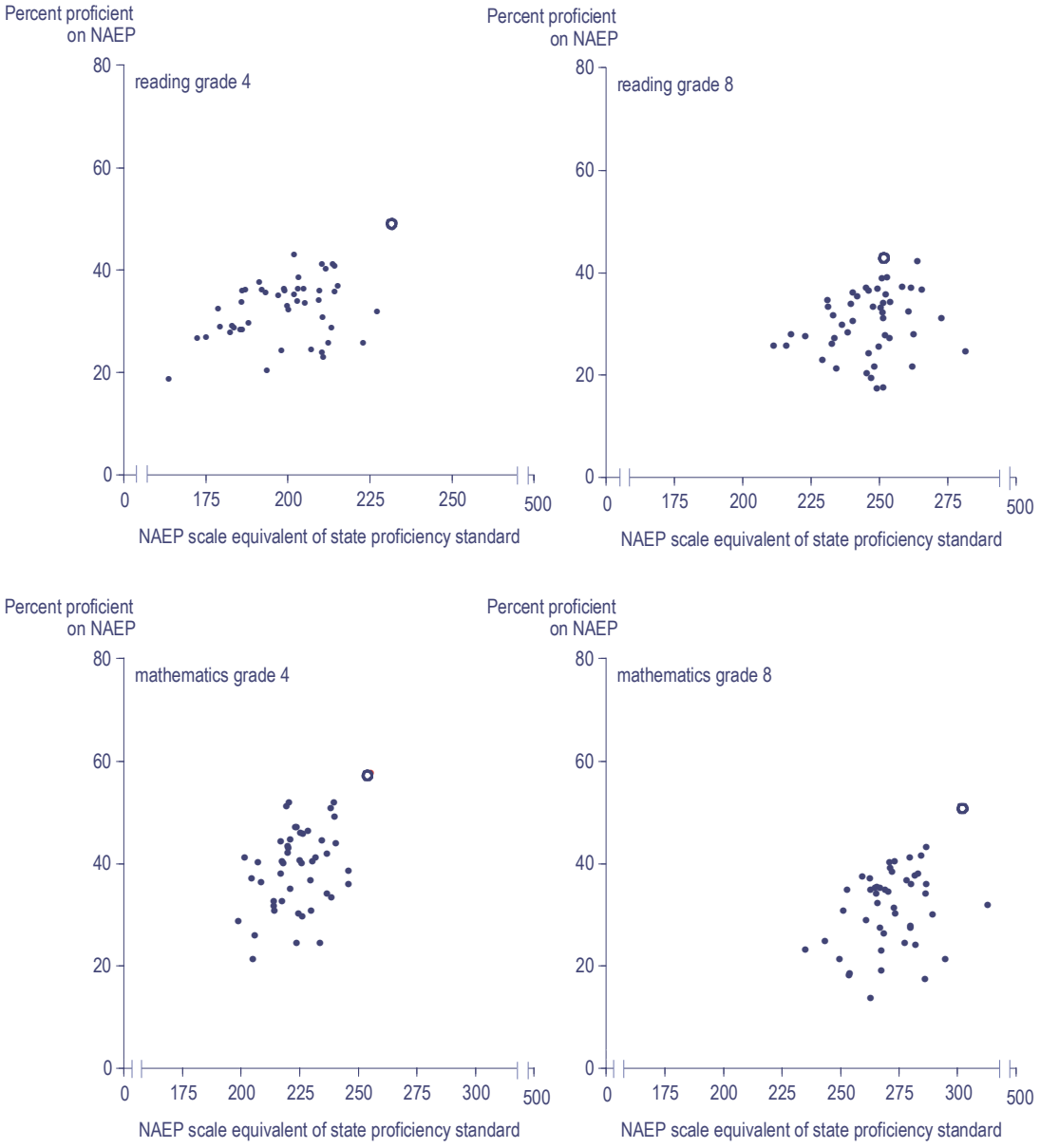
* Statistically significant at $p < .05$.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading and Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Whereas table 5 addresses the question of how the variability of performance standards relates to the percentages of students meeting the standards, figure 4 and table 6 address the question of how the variation among performance standards relates to the performance of students on NAEP. Figure 4 displays, for each subject and grade, the percentage of each state's students meeting the NAEP *Proficient* standard as a function of the placement of their own standard for *proficient* performance. Table 6 summarizes the linear relationships. Although three of the functions slope upward, this is mainly caused by a single state that set a high standard and had high scores. If that state is removed (the circled dot on figure 4), the squared correlations are .10 (from .16) for grade 4 reading, .04 (unchanged) for grade 8 reading, .09 (from .15) for grade 4 mathematics, and .06 (from .12) from grade 8 mathematics. The two squared correlations for grade 4 are statistically significant, but the two grade 8 relationships are not.

In general, from figure 4, we see that setting a higher state standard is not necessarily associated with higher performance on NAEP. In grade 8 at least, students in states with high standards for *proficient* performance score just about the same on NAEP as students in states with low standards for proficiency.

Figure 4. Relationship between the percentage of students scoring proficient on NAEP and the difficulty of grades 4 and 8 state standards for reading and mathematics as measured by the state's respective NAEP scale equivalent: 2007



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading and Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table 6. Relationship between the percentage of students scoring proficient on NAEP and the difficulty of grades 4 and 8 state standards as measured by the state's respective NAEP scale equivalent, by subject: 2007

Percent proficient on NAEP = f(state standards as measured by the state's NAEP scale equivalent)							
		Grade 4			Grade 8		
Subject		**Intercept**	**Slope**	**R^2**	**Intercept**	**Slope**	**R^2**
Reading	Estimate	-1.3	.2 *	.16	6.8	.1	.04
	Standard error	*11.36*	*.06*	†	*16.12*	*.07*	†
Mathematics	Estimate	-17.3	.3 *	.15	-18.9	.2 *	.12
	Standard error	*19.92*	*.09*	†	*20.26*	*.07*	†

† Not applicable.

* Statistically significant at $p < .05$.

NOTE: Removing one state that set a high standard and had high scores, the R^2 are .10 (from .16) for grade 4 reading, .04 (unchanged) for grade 8 reading, .09 (from .15) for mathematics grade 4, and .06 (from .12) from mathematics grade 8. The two R^2 for grade 4 are statistically significant, but the two grade 8 relationships are not.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading and Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

4 COMPARING 2007 WITH 2005 STATE PERFORMANCE STANDARDS

Although the NAEP assessment in reading and mathematics did not change between 2005 and 2007, some states made changes in their state assessments in these subjects during the same period, changes substantial enough that these states indicated that their 2005 scores were not comparable to their 2007 scores.[23] Nevertheless, both 2005 and 2007 scores could be mapped onto the NAEP scale as a means for comparison. For these states, the analyses compared the NAEP equivalent scores estimated for 2007 with those for 2005. Significant differences in NAEP scale equivalents might reflect changes in policies and/or practices that occurred between the years in addition to the changes in state assessments and standards.

Other states reported no changes in their state assessments in the same period and indicated that their 2005 scores were comparable to their 2007 scores. For these states, the analyses compared the NAEP equivalent scores estimated for 2007 to those for 2005 to evaluate the stability of the mapping of each state's standard for proficient performance onto the NAEP scale.

When the 2005 and 2007 NAEP equivalents of the state standards are not stable, that is, the NAEP equivalent score for 2007 is statistically significantly different from that of the 2005, further investigation is warranted. Several factors could lead to such instability. For example, changes in classroom instructional practices or curricula might have placed more emphasis on subject matter covered more on the state test than on NAEP from one assessment year to the next, or 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.[24]

Regardless of whether states reported that 2005 scores are comparable to 2007 or not, when NAEP scale equivalents are significantly different, further investigations can help ascertain the factors that may have contributed to the differences in the NAEP scale equivalents of state standards seen in this study. When the 2005 NAEP equivalents of the state standards are not different from those for 2007, that is, when standards are considered stable, NAEP can be used to corroborate the state reported progress (or lack of progress) through further analysis, an issue discussed in Section 5.

This section makes comparisons between the 2005 and 2007 mappings in reading and mathematics for grades 4 and 8. The 2005 mappings in this report will not necessarily match previously published results (U.S. Department of Education 2007). Methodological differences between the procedures used in the two analyses may result in small differences.[25] Moreover, since the release of the 2005 mapping study, some states have revised their 2005 assessment data files and other states have made public previously unavailable results.

---

[23] This was reported in a survey conducted for this study to gain contextual information about the general characteristics of state assessment programs and, specifically, to help identify changes in states' assessments between the 2004-05 and 2006-07 school years that could affect the interpretation of the mapping results. See appendix B for more information on the survey.

[24] These issues were not covered by the survey of state assessment programs referenced above.

[25] The small differences are not large enough to change the whole number scale value reported as the NAEP equivalent.

Reading

Table 7 displays the availability of state assessment data in 2005 and 2007 suitable for implementing the mapping of the states' grades 4 and 8 reading standards onto the NAEP scale. Table 7 also shows, for each grade, whether changes in the states' assessments between 2005 and 2007 were deemed by state representatives to affect the comparability of the 2005 with the 2007 reported results.[26] States with both years of data are listed in table 8 by grade and by whether those data are comparable. In grade 4 reading, of the 34 states with valid test data in both years, 22 states indicated that no significant changes in their tests were made that would affect the comparability of test results across the two years. For grade 8 reading, of the 38 states with valid test data in both years, 14 indicated that their scores were not comparable and 24 indicated comparability of results.

For states with both years of data, tables 9 and 10 display, for each year, the number of public schools selected for NAEP in each state, the percentage of these schools included in the analyses, and the percentage of the student population represented by the schools.

Tables 11 and 12 compare the NAEP scale equivalents between the two years for grades 4 and 8, respectively, according to whether states reported comparable assessment results. Table 11 shows that, for the 12 states indicating substantive changes in their grade 4 reading assessments, 8 showed significant differences between the 2005 and 2007 estimates of the NAEP equivalents of their state standards. Half of these showed an increase of up to 12 points (Idaho), and half showed a decrease of up to 24 points (Wyoming). Table 11 also shows that, among the 22 states indicating no substantive changes in grade 4 state tests, 14 states did not show statistically significant differences between their NAEP scale equivalents in 2005 and 2007; 8 states showed statistically significant differences in the estimated NAEP scale equivalent, with 5 showing standards that are as much as 11 points higher (New Jersey) and 3 showing a decrease of up to 6 points (South Carolina).

Table 12 shows that among those states indicating substantive changes in their grade 8 reading assessments, seven showed significant differences between the 2005 and 2007 estimates of the NAEP equivalents of their state standards; all seven showed lower 2007 NAEP scale equivalent of their standards, by up to 31 points (Wyoming).

Table 12 also shows that, among the 24 states indicating no changes in their state tests, the NAEP equivalent standards of 13 states in 2007 were not statistically different from their standards in 2005. The 11 remaining states showed statistically significant differences in the estimates of the NAEP scale equivalent, 8 of which showed decreases in NAEP scale equivalent of state standards of up to 12 points (Pennsylvania) and 3 showed increases in NAEP equivalent of state standards of up to 5 points (Maryland).

[26] Tables B-1 to B-3 of appendix B summarize for each state selected changes to the main state assessment in reading and mathematics between 2005 and 2007 and information about the comparability of the reported results between 2005 and 2007.

Table 7. State assessment data availability and state reports of whether 2005 and 2007 assessment results are comparable in grades 4 and 8 reading, by state: 2005 and 2007

State/jurisdiction	Grade 4			Grade 8		
	2005 data	2007 data	Comparable results	2005 data	2007 data	Comparable results
Alabama	√	√	Yes	√	√	Yes
Alaska	√	√	Yes	√	√	Yes
Arizona	—	√	Yes	√	√	Yes
Arkansas	√	√	Yes	√	√	Yes
California	√	√	Yes	√	√	Yes
Colorado	√	√	Yes	√	√	Yes
Connecticut	√	√	No	√	√	No
Delaware	—	√	No	√	√	No
District of Columbia	—	—	No	—	—	No
Florida	√	√	Yes	√	√	Yes
Georgia	√	√	No	√	√	No
Hawaii	√	√	No	√	√	No
Idaho	√	√	No	√	√	No
Illinois	—	√	No	√	√	Yes
Indiana	√	√	Yes	√	√	Yes
Iowa	√	√	Yes	√	√	Yes
Kansas	—	√	No	√	√	No
Kentucky	√	√	No	—	√	No
Louisiana	√	√	Yes	√	√	Yes
Maine	√	√	No	√	√	No
Maryland	√	√	Yes	√	√	Yes
Massachusetts	√	√	Yes	—	√	No
Michigan	√	√	No	—	√	No
Minnesota	—	√	No	—	√	No
Mississippi	√	√	Yes	√	√	Yes
Missouri	—	√	No	—	√	No
Montana	√	√	No	√	√	No
Nebraska	—	—	No	—	—	No
Nevada	—	√	No	√	√	Yes
New Hampshire	—	√	No	—	√	No
New Jersey	√	√	Yes	√	√	Yes
New Mexico	√	√	Yes	√	√	Yes
New York	√	√	No	√	√	No
North Carolina	√	√	Yes	√	√	Yes
North Dakota	√	√	Yes	√	√	Yes
Ohio	√	√	Yes	√	√	Yes
Oklahoma	√	√	No	√	√	No
Oregon	—	√	No	√	√	No
Pennsylvania	—	√	No	√	√	Yes
Rhode Island	—	√	Yes	—	√	Yes
South Carolina	√	√	Yes	√	√	Yes
South Dakota	—	√	Yes	—	√	Yes
Tennessee	√	√	Yes	√	√	Yes
Texas	√	√	Yes	√	√	Yes
Utah	—	—	Yes	—	—	No
Vermont	—	√	Yes	—	√	Yes
Virginia	—	√	No	√	√	No
Washington	√	√	Yes	—	√	No
West Virginia	√	√	No	√	√	No
Wisconsin	√	√	Yes	√	√	Yes
Wyoming	√	√	No	√	√	No

√ State assessment data available.

— State assessment data not available.

SOURCE: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2007 Survey of State Assessment Program Characteristics.

Table 8. States with both 2005 and 2007 data suitable to implement the mapping of grades 4 and 8 state reading standards, by whether the reported results are directly comparable

2005 and 2007 state assessment reported grade 4 results directly comparable	2005 and 2007 state assessment reported grade 8 results directly comparable
Alabama	Alabama
Alaska	Alaska
Arkansas	Arizona
California	Arkansas
Colorado	California
Florida	Colorado
Indiana	Florida
Iowa	Illinois
Louisiana	Indiana
Maryland	Iowa
Massachusetts	Louisiana
Mississippi	Maryland
New Jersey	Mississippi
New Mexico	Nevada
North Carolina	New Jersey
North Dakota	New Mexico
Ohio	North Carolina
South Carolina	North Dakota
Tennessee	Ohio
Texas	Pennsylvania
Washington	South Carolina
Wisconsin	Tennessee
	Texas
	Wisconsin
2005 and 2007 state assessment reported grade 4 results not comparable	**2005 and 2007 state assessment reported grade 8 results not comparable**
Connecticut	Connecticut
Georgia	Delaware
Hawaii	Georgia
Idaho	Hawaii
Kentucky	Idaho
Maine	Kansas
Michigan	Maine
Montana	Montana
New York	New York
Oklahoma	Oklahoma
West Virginia	Oregon
Wyoming	Virginia
	West Virginia
	Wyoming

SOURCE: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2007 Survey of State Assessment Program Characteristics.

Table 9. Number of NAEP schools, percentage of NAEP schools available for comparing state assessment results with NAEP results in grade 4 reading, and percentage of the student population represented in these comparison schools, by state: 2005 and 2007

State/jurisdiction	2005 NAEP schools[1]	2005 Percent of NAEP schools matched	2005 Percent of population represented	2007 NAEP schools[1]	2007 Percent of NAEP schools matched	2007 Percent of population represented
Alabama	130	98.5	97.6	110	99.1	99.1
Alaska[2]	160	61.8	83.7	180	99.4	99.9
Arkansas	150	84.8	91.5	120	96.6	97.6
California	450	94.6	96.3	320	97.8	99.0
Colorado	150	91.8	97.1	120	95.8	99.0
Connecticut	130	100.0	100.0	110	100.0	100.0
Florida	170	94.1	96.3	160	97.6	97.2
Georgia	180	92.6	91.7	160	98.7	96.4
Hawaii	130	100.0	100.0	120	99.1	99.1
Idaho	160	95.5	94.9	130	95.5	91.2
Indiana	140	100.0	100.0	110	100.0	100.0
Iowa	130	96.2	97.0	140	97.8	96.7
Kentucky	150	99.3	99.1	120	97.4	98.1
Louisiana	140	99.3	98.5	110	97.2	98.4
Maine[2]	190	74.1	81.3	150	93.4	95.4
Maryland	130	98.4	99.2	110	98.2	98.4
Massachusetts	200	98.5	99.7	170	100.0	100.0
Michigan	140	92.3	95.3	120	99.2	98.7
Mississippi	130	99.2	99.8	120	97.4	97.1
Montana	240	80.5	94.3	190	98.9	99.1
New Jersey	140	99.3	98.9	110	98.2	95.1
New Mexico[2]	160	83.9	83.9	130	95.3	97.9
New York	190	97.9	98.8	150	99.3	99.8
North Carolina	180	96.0	97.4	170	97.6	96.5
North Dakota	260	74.3	93.0	210	80.5	93.3
Ohio	200	98.5	99.3	160	98.1	99.3
Oklahoma	180	99.4	99.8	140	98.5	98.8
South Carolina	120	99.2	99.3	110	97.2	98.7
Tennessee	140	98.6	97.8	120	100.0	100.0
Texas	380	98.2	97.6	300	98.6	97.9
Washington	140	97.8	99.0	130	99.2	100.0
West Virginia[2]	200	97.4	97.9	150	92.5	89.7
Wisconsin[2]	170	58.6	65.3	130	65.4	71.0
Wyoming	170	85.9	96.6	170	96.5	97.2

[1] Rounded to the nearest 10 for confidentiality.

[2] The percentage of the student population represented by the NAEP schools used in the estimations was less than 90 percent in at least one of the years.

NOTE: In the comparison schools, the population represented by NAEP is less than 100 percent of the total population where state assessment scores are missing for some schools. Scores may be missing either because of the failure to match schools in the two surveys or the suppression of scores where there are too few students.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table 10. Number of NAEP schools, percentage of NAEP schools available for comparing state assessment results with NAEP results in grade 8 reading, and percentage of the student population represented in these comparison schools, by state: 2005 and 2007

State/jurisdiction	2005			2007		
	NAEP schools[1]	Percent of NAEP schools matched	Percent of population represented	NAEP schools[1]	Percent of NAEP schools matched	Percent of population represented
Alabama	110	98.2	98.1	120	100.0	100.0
Alaska[2]	100	52.9	89.1	110	98.2	99.3
Arizona	130	96.2	99.1	130	97.7	99.2
Arkansas[2]	130	84.0	89.0	120	91.1	94.5
California	370	95.2	97.2	310	97.1	99.0
Colorado	120	90.0	98.2	120	93.1	98.5
Connecticut	110	96.2	97.0	100	100.0	100.0
Delaware	40	86.0	92.9	50	97.8	100.0
Florida	160	96.3	95.2	160	98.7	98.6
Georgia	120	92.7	91.9	120	97.5	95.6
Hawaii	70	98.5	99.9	70	100.0	100.0
Idaho	100	94.1	97.1	110	97.2	99.0
Illinois	190	98.4	98.2	200	98.0	99.3
Indiana	110	98.1	97.9	110	100.0	100.0
Iowa	110	98.2	97.0	130	97.0	96.8
Kansas	120	97.4	99.1	150	97.3	98.0
Louisiana	110	98.2	98.5	110	96.4	97.7
Maine[2]	130	67.7	80.2	130	94.7	97.4
Maryland	110	98.1	99.2	110	99.1	97.3
Mississippi	120	96.5	97.0	110	97.4	97.9
Montana	160	81.8	96.3	170	98.2	99.4
Nevada	80	87.2	92.9	70	93.2	93.3
New Jersey	110	99.1	96.9	110	100.0	100.0
New Mexico[2]	110	81.1	84.7	110	97.3	99.4
New York	180	95.1	95.3	160	98.1	98.5
North Carolina	140	95.0	97.5	150	99.3	99.8
North Dakota	180	73.6	92.9	190	70.3	90.0
Ohio	140	95.1	96.9	190	98.4	99.1
Oklahoma	150	96.6	97.1	150	96.6	96.8
Oregon	120	99.2	99.8	110	96.5	99.0
Pennsylvania	110	94.5	96.0	110	98.2	97.6
South Carolina	110	96.3	95.6	110	97.2	98.5
Tennessee	110	99.1	99.5	120	99.2	99.1
Texas	280	97.1	98.1	220	96.4	97.4
Virginia	110	100.0	100.0	110	99.1	98.9
West Virginia	110	97.3	98.8	120	91.5	91.1
Wisconsin[2]	120	79.7	86.1	130	74.6	82.1
Wyoming	80	98.7	96.8	80	95.1	96.1

[1] Rounded to the nearest 10 for confidentiality.

[2] The percentage of the student population represented by the NAEP schools used in the estimations was less than 90 percent in at least one of the years.

NOTE: In the comparison schools, the population represented by NAEP is less than 100 percent of the total population where state assessment scores are missing for some schools. Scores may be missing either because of the failure to match schools in the two surveys or the suppression of scores where there are too few students.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table 11. Difference between the estimated NAEP scale equivalents of state grade 4 reading proficient standards and their standard error, by state: 2005 and 2007

State/jurisdiction	2005 NAEP scale equivalent	2005 Standard error	2007 NAEP scale equivalent	2007 Standard error	Difference 2007-2005	Standard error
2005 and 2007 state assessment reported results are comparable						
Alabama	172	2.1	179	1.5	6.8 *	2.63
Alaska	182	2.8	183	0.9	1.1	2.93
Arkansas	217	1.4	213	1.4	-4.1 *	1.96
California	210	0.7	210	0.9	0.3	1.10
Colorado	186	1.5	187	1.5	0.5	2.08
Florida	202	0.9	209	0.8	6.8 *	1.22
Indiana	199	1.2	199	1.3	0.4	1.74
Iowa	197	1.4	199	1.7	1.8	2.21
Louisiana	198	1.4	193	2.2	-4.5	2.57
Maryland	187	1.5	186	1.5	-1.0	2.15
Massachusetts	234	0.9	232	1.2	-2.3	1.54
Mississippi	161	2.1	163	1.3	2.5	2.48
New Jersey	191	1.7	201	2.0	10.6 *	2.61
New Mexico	208	1.0	210	0.7	1.6	1.23
North Carolina	183	1.3	183	1.0	-0.8	1.64
North Dakota	204	0.7	201	1.0	-2.5 *	1.21
Ohio	199	1.8	198	2.2	-0.5	2.88
South Carolina	228	1.1	223	1.5	-5.9 *	1.86
Tennessee	170	1.5	175	1.7	4.9 *	2.31
Texas	190	1.0	188	1.6	-2.8	1.85
Washington	197	1.9	203	2.1	5.9 *	2.80
Wisconsin	189	1.7	193	2.0	4.1	2.60
2005 and 2007 state assessment reported results are not comparable						
Connecticut	212	1.1	213	1.6	0.8	1.95
Georgia	174	1.6	185	1.3	11.0 *	2.06
Hawaii	205	0.8	212	1.0	7.2 *	1.28
Idaho	185	3.2	197	1.4	11.9 *	3.43
Kentucky	206	1.6	205	1.6	-1.6	2.24
Maine	224	1.1	214	1.0	-10.1 *	1.55
Michigan	182	3.8	178	2.5	-4.1	4.57
Montana	197	1.5	203	1.2	5.6 *	1.93
New York	207	1.2	209	1.4	2.6	1.83
Oklahoma	182	2.3	172	3.7	-10.3 *	4.38
West Virginia	186	1.3	182	1.4	-4.1 *	1.92
Wyoming	228	0.6	204	1.2	-23.8 *	1.30

* Difference is statistically significant at $p < .05$.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table 12. Difference between the estimated NAEP scale equivalents of state grade 8 reading proficient standards and their standard error, by state: 2005 and 2007

State/jurisdiction	2005 NAEP scale equivalent	2005 Standard error	2007 NAEP scale equivalent	2007 Standard error	Difference 2007-2005	Standard error
2005 and 2007 state assessment reported results are comparable						
Alabama	236	1.1	234	1.5	-2.7	1.92
Alaska	230	1.3	233	1.9	2.7	2.25
Arizona	244	1.1	245	1.1	1.0	1.58
Arkansas	254	1.0	249	1.4	-5.2 *	1.67
California	262	0.7	261	0.6	-0.7	0.93
Colorado	229	1.9	230	1.4	1.9	2.33
Florida	265	1.2	262	0.8	-3.0 *	1.45
Illinois	245	1.1	236	1.5	-9.6 *	1.89
Indiana	249	1.9	251	0.7	1.3	2.04
Iowa	250	1.0	252	1.1	1.4	1.45
Louisiana	251	1.2	246	1.3	-4.7 *	1.81
Maryland	245	1.7	250	1.2	5.0 *	2.09
Mississippi	246	1.4	251	0.6	4.5 *	1.52
Nevada	253	0.9	247	1.0	-5.2 *	1.38
New Jersey	250	1.2	252	1.1	1.8	1.67
New Mexico	251	1.4	248	1.0	-2.1	1.74
North Carolina	217	1.4	217	1.2	0.4	1.82
North Dakota	255	0.8	251	1.4	-4.0 *	1.62
Ohio	241	1.6	240	1.9	-1.0	2.52
Pennsylvania	258	1.7	245	1.4	-12.3 *	2.25
South Carolina	276	1.2	281	1.0	4.8 *	1.55
Tennessee	221	1.8	211	2.5	-10.6 *	3.09
Texas	225	0.9	222	1.1	-2.6	1.41
Wisconsin	229	1.5	231	1.4	1.5	2.06
2005 and 2007 state assessment reported results are not comparable						
Connecticut	242	1.4	245	1.1	2.6	1.79
Delaware	242	1.1	240	1.0	-2.3	1.50
Georgia	224	1.3	215	1.7	-8.4 *	2.17
Hawaii	261	1.2	245	0.7	-16.7 *	1.37
Idaho	235	1.9	233	1.0	-2.5	2.18
Kansas	242	1.4	241	1.0	-1.3	1.68
Maine	275	1.3	261	0.9	-14.4 *	1.62
Montana	253	0.9	250	1.5	-2.7	1.79
New York	268	1.3	260	0.9	-7.9 *	1.58
Oklahoma	244	1.3	232	1.6	-11.7 *	2.08
Oregon	254	1.3	251	1.2	-3.1	1.76
Virginia	243	1.3	239	1.2	-4.3 *	1.83
West Virginia	228	1.8	229	1.3	0.2	2.22
Wyoming	278	1.4	247	1.1	-31.2 *	1.77

* Difference is statistically significant at $p < .05$.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Mathematics

Table 13 displays the availability of state assessment data in 2005 and 2007 suitable for implementing the mapping of grades 4 and 8 mathematics standards. It also displays, for each grade, whether changes in the states' assessments between 2005 and 2007 were deemed to affect the direct comparability of the 2005 and 2007 reported results. States with both years of data are listed in table 14 by grade and by whether those data are comparable according to state assessment staff. In grade 4 mathematics, of the 35 states with valid test data in both years, 14 indicated that their 2005 scores were not comparable to their 2007 scores and 21 states indicated that no significant changes in their tests were made. For grade 8 mathematics, of the 39 states with valid test data in both years, 18 indicated that their scores were not comparable and 21 indicated comparability of results.

For states with both years of data, tables 15 and 16 display, for each year, the number of public schools selected for NAEP in each state, the percentage of these schools included in the analyses, and the percentage of the student population represented by the schools.

Tables 17 and 18 compare the NAEP scale equivalent between the two years for grades 4 and 8, respectively, according to whether states reported comparable assessment results. Table 17 shows that for the 14 states indicating substantive changes in their grade 4 assessments, 11 showed significant differences between the 2005 and 2007 NAEP equivalents of their state standards. Six of them had lower 2007 NAEP equivalent of state standards with decreases of up to 34 points (Wyoming), and five had higher 2007 NAEP equivalent standards, with increases of up to 28 points (North Carolina).

Table 17 also shows that among the 21 states indicating no substantive changes in grade 4 state tests, 15 did not have statistically significant differences between their NAEP scale equivalents in 2005 and 2007. Six states had statistically significant differences in the NAEP scale equivalent, with two showing increases of up to 4 points (Washington), and four showing decreases of up to 8 points (Maryland).

Table 18 shows that among those 18 states indicating substantive changes in their grade 8 mathematics assessments, 12 showed significant differences between the 2005 and 2007 estimates of the NAEP equivalents of their state standards: 9 states showed lower 2007 NAEP equivalent standards, by up to 25 points (Illinois), and 3 showed increases of up to 23 points (North Carolina). Table 18 also shows that, among the 21 states indicating no changes in their state tests, the NAEP scale equivalent of state standards of 14 states in 2007 were not statistically different from the standards in 2005. The remaining seven states had statistically significant differences in their NAEP equivalent standards; six showed decreases by up to 12 points (Georgia), and South Carolina increased its NAEP equivalent standard by 7 points.

Such discrepancies illustrate that the method used for mapping state standards onto the NAEP scales may produce an apparent change in the state's standard, causing it to appear somewhat easier or more stringent. For this reason, the results of studies like this one need to be re-estimated with each NAEP state assessment to ensure that the NAEP-equivalent mapping is up-to-date. This method relies on NAEP and state tests to track the same progress over time. Section 5 explores this issue in more detail.

Table 13. State assessment data availability and state reports of whether 2005 and 2007 assessment results are comparable in grades 4 and 8 mathematics, by state: 2005 and 2007

State/jurisdiction	Grade 4			Grade 8		
	2005 data	2007 data	Comparable results	2005 data	2007 data	Comparable results
Alabama	√	√	Yes	—	√	Yes
Alaska	√	√	Yes	√	√	Yes
Arizona	—	√	Yes	√	√	Yes
Arkansas	√	√	Yes	√	√	Yes
California	√	√	Yes	—	—	Yes
Colorado	√	√	Yes	√	√	Yes
Connecticut	√	√	No	√	√	No
Delaware	—	√	No	√	√	No
District of Columbia	—	—	No	—	—	No
Florida	√	√	Yes	√	√	Yes
Georgia	√	√	Yes	√	√	Yes
Hawaii	√	√	No	√	√	No
Idaho	√	√	No	√	√	No
Illinois	—	√	No	√	√	No
Indiana	√	√	Yes	√	√	Yes
Iowa	√	√	Yes	√	√	Yes
Kansas	√	√	No	—	√	No
Kentucky	—	√	No	√	√	No
Louisiana	√	√	Yes	√	√	Yes
Maine	√	√	No	√	√	No
Maryland	√	√	Yes	√	√	Yes
Massachusetts	√	√	Yes	√	√	No
Michigan	√	√	No	√	√	No
Minnesota	—	√	No	—	√	No
Mississippi	√	√	Yes	√	√	Yes
Missouri	√	√	No	√	√	No
Montana	√	√	No	√	√	No
Nebraska	—	—	No	—	—	No
Nevada	—	√	No	√	√	Yes
New Hampshire	—	√	No	—	√	No
New Jersey	√	√	Yes	√	√	Yes
New Mexico	√	√	Yes	√	√	Yes
New York	√	√	No	√	√	No
North Carolina	√	√	No	√	√	No
North Dakota	√	√	Yes	√	√	Yes
Ohio	√	√	No	√	√	Yes
Oklahoma	√	√	No	√	√	No
Oregon	—	√	No	√	√	No
Pennsylvania	—	√	No	√	√	Yes
Rhode Island	—	√	Yes	—	√	Yes
South Carolina	√	√	Yes	√	√	Yes
South Dakota	—	√	Yes	—	√	Yes
Tennessee	√	√	Yes	√	√	Yes
Texas	√	√	Yes	√	√	Yes
Utah	—	—	Yes	—	—	Yes
Vermont	—	√	Yes	—	√	Yes
Virginia	—	√	No	√	√	No
Washington	√	√	Yes	—	√	No
West Virginia	√	√	No	√	√	No
Wisconsin	√	√	Yes	√	√	Yes
Wyoming	√	√	No	√	√	No

√ State assessment data available.

— State assessment data not available.

SOURCE: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2007 Survey of State Assessment Program Characteristics.

Table 14. States with both 2005 and 2007 data suitable to implement the mapping of grades 4 and 8 mathematics standards, by whether the reported results are directly comparable

2005 and 2007 state assessment reported grade 4 results directly comparable	2005 and 2007 state assessment reported grade 8 results directly comparable
Alabama	Alaska
Alaska	Arizona
Arkansas	Arkansas
California	Colorado
Colorado	Florida
Florida	Georgia
Georgia	Indiana
Indiana	Iowa
Iowa	Louisiana
Louisiana	Maryland
Maryland	Mississippi
Massachusetts	Nevada
Mississippi	New Jersey
New Jersey	New Mexico
New Mexico	North Dakota
North Dakota	Ohio
South Carolina	Pennsylvania
Tennessee	South Carolina
Texas	Tennessee
Washington	Texas
Wisconsin	Wisconsin
2005 and 2007 state assessment reported grade 4 results not comparable	**2005 and 2007 state assessment reported grade 8 results not comparable**
Connecticut	Connecticut
Hawaii	Delaware
Idaho	Hawaii
Kansas	Idaho
Maine	Illinois
Michigan	Kentucky
Missouri	Maine
Montana	Massachusetts
New York	Michigan
North Carolina	Missouri
Ohio	Montana
Oklahoma	New York
West Virginia	North Carolina
Wyoming	Oklahoma
	Oregon
	Virginia
	West Virginia
	Wyoming

SOURCE: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2007 Survey of State Assessment Program Characteristics.

Table 15. Number of NAEP schools, percentage of NAEP schools available for comparing state assessment results with NAEP results in grade 4 mathematics, and percentage of the student population in these comparison schools, by state: 2005 and 2007

State/jurisdiction	2005 NAEP schools[1]	2005 Percent of NAEP schools matched	2005 Percent of population represented	2007 NAEP schools[1]	2007 Percent of NAEP schools matched	2007 Percent of population represented
Alabama	130	98.5	97.9	110	99.1	99.1
Alaska	150	70.6	91.2	180	100.0	100.0
Arkansas	150	84.8	91.9	120	96.6	97.5
California	450	94.4	96.4	330	97.5	98.9
Colorado	150	92.5	96.9	120	95.8	99.1
Connecticut	130	100.0	100.0	110	100.0	100.0
Florida	170	94.1	96.6	160	97.6	97.2
Georgia	180	92.6	92.1	160	98.7	96.4
Hawaii	130	100.0	100.0	120	99.1	99.0
Idaho	160	95.6	95.1	130	95.5	91.8
Indiana	140	100.0	100.0	110	100.0	100.0
Iowa	130	95.4	96.2	140	97.8	96.9
Kansas	140	96.4	98.0	140	98.6	99.0
Louisiana	140	99.3	98.3	110	97.2	98.4
Maine[2]	190	74.2	82.2	150	93.4	95.6
Maryland	130	99.2	99.7	110	98.2	98.5
Massachusetts	200	99.0	99.8	170	100.0	100.0
Michigan	140	92.9	95.5	120	99.2	98.8
Mississippi	130	100.0	100.0	120	97.4	97.2
Missouri	160	97.5	98.7	130	98.4	99.6
Montana	250	77.9	93.3	190	98.9	99.3
New Jersey	140	99.3	98.6	110	98.2	95.0
New Mexico[2]	160	83.3	84.7	130	93.8	97.5
New York	190	97.9	98.9	150	99.3	99.8
North Carolina	180	96.0	97.5	170	97.6	96.4
North Dakota	260	74.3	93.3	210	81.3	93.1
Ohio	200	99.0	99.4	160	98.1	99.4
Oklahoma	180	98.9	99.6	140	98.6	98.7
South Carolina	120	99.2	99.2	110	97.2	98.3
Tennessee	140	98.6	98.2	120	100.0	100.0
Texas	380	98.4	97.7	300	98.6	98.0
Washington	140	97.8	99.0	130	99.2	100.0
West Virginia	200	97.4	98.0	150	92.5	89.5
Wisconsin[2]	170	58.6	65.5	130	65.4	70.7
Wyoming	160	89.0	97.2	170	97.6	97.2

[1] Rounded to the nearest 10 for confidentiality.

[2] The percentage of the student population represented by the NAEP schools used in the estimations was less than 90 percent in at least one of the years.

NOTE: In the comparison schools, the population represented by NAEP is less than 100 percent of the total population where state assessment scores are missing for some schools. Scores may be missing either because of the failure to match schools in the two surveys or the suppression of scores where there are too few students.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table 16. Number of NAEP schools, percentage of NAEP schools available for comparing state assessment results with NAEP results in grade 8 mathematics, and percentage of the student population in the comparison schools, by state: 2005 and 2007

State/jurisdiction	2005 NAEP schools[1]	2005 Percent of NAEP schools matched	2005 Percent of population represented	2007 NAEP schools[1]	2007 Percent of NAEP schools matched	2007 Percent of population represented
Alaska	100	58.4	90.7	110	98.2	99.3
Arizona	130	96.2	98.7	130	97.7	99.1
Arkansas[2]	130	84.0	88.8	130	90.4	94.5
Colorado	120	89.3	97.7	120	93.1	98.6
Connecticut	110	96.2	96.7	100	100.0	100.0
Delaware	40	86.0	93.4	50	100.0	100.0
Florida	160	95.7	95.7	160	98.7	98.6
Georgia	120	92.7	92.0	120	97.5	95.4
Hawaii	70	98.5	99.8	70	95.7	99.8
Idaho	100	93.2	97.3	100	98.1	99.1
Illinois	190	98.4	98.6	200	98.0	99.4
Indiana	110	98.1	98.1	110	100.0	100.0
Iowa	110	98.2	96.6	140	96.3	96.9
Kentucky	120	99.1	99.2	110	98.2	98.7
Louisiana	110	98.2	98.5	110	96.4	97.7
Maine[2]	130	67.2	80.5	130	94.7	97.6
Maryland	110	98.1	99.2	110	99.1	97.2
Massachusetts	130	97.7	99.4	130	99.3	99.2
Michigan	120	95.7	97.6	120	96.7	97.9
Mississippi	120	96.5	97.6	110	97.4	97.7
Missouri	130	96.2	97.7	130	94.7	96.2
Montana	160	79.9	96.0	170	98.2	99.4
Nevada	80	88.3	92.5	80	93.3	93.6
New Jersey	110	99.1	96.9	110	100.0	100.0
New Mexico[2]	110	81.1	84.2	110	97.3	99.6
New York	180	95.1	95.7	160	98.1	98.5
North Carolina	140	95.0	97.7	150	99.3	99.7
North Dakota[2]	180	73.4	92.5	180	70.3	89.6
Ohio	140	95.1	97.0	190	98.9	98.8
Oklahoma	150	95.9	97.2	150	96.6	96.8
Oregon	120	99.2	99.8	110	96.5	99.2
Pennsylvania	110	94.5	96.1	110	98.2	97.5
South Carolina	110	97.2	95.8	110	97.2	98.8
Tennessee	110	99.1	99.4	120	99.2	99.2
Texas	280	97.1	98.0	220	96.4	97.6
Virginia	110	100.0	100.0	110	100.0	100.0
West Virginia	110	97.3	99.0	120	91.5	91.0
Wisconsin[2]	120	79.7	86.5	130	74.6	82.6
Wyoming	80	96.3	96.5	80	96.3	97.1

[1] Rounded to the nearest 10 for confidentiality.

[2] The percentage of the student population represented by the NAEP schools used in the estimations was less than 90 percent in at least one of the years.

NOTE: In the comparison schools, the population represented by NAEP is less than 100 percent of the total population where state assessment scores are missing for some schools. Scores may be missing either because of the failure to match schools in the two surveys or the suppression of scores where there are too few students.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table 17. Difference between the estimated NAEP scale equivalents of state grade 4 mathematics proficient standards and their standard error, by state: 2005 and 2007

State/jurisdiction	2005 NAEP scale equivalent	2005 Standard error	2007 NAEP scale equivalent	2007 Standard error	Difference 2007-2005	Standard error
2005 and 2007 state assessment reported results are comparable						
Alabama	207	0.9	205	1.5	-1.6	1.72
Alaska	222	1.1	216	1.3	-5.8 *	1.74
Arkansas	236	1.1	229	0.6	-6.7 *	1.26
California	231	0.6	226	0.7	-5.1 *	0.92
Colorado	201	1.2	201	1.6	-0.1	2.04
Florida	230	0.8	230	0.8	-0.7	1.19
Georgia	215	1.0	213	0.8	-1.4	1.28
Indiana	225	0.7	228	0.9	2.5 *	1.14
Iowa	219	0.8	220	1.1	0.4	1.38
Louisiana	223	0.9	223	1.3	0.2	1.61
Maryland	215	1.1	206	1.3	-8.3 *	1.65
Massachusetts	255	0.8	254	1.0	-0.9	1.24
Mississippi	206	1.1	204	0.8	-1.6	1.38
New Jersey	221	1.4	220	1.1	-0.9	1.77
New Mexico	232	1.3	233	0.8	0.4	1.50
North Dakota	224	0.8	226	1.0	1.8	1.29
South Carolina	246	1.0	245	0.9	-1.4	1.33
Tennessee	200	1.2	198	1.3	-1.4	1.75
Texas	219	1.0	217	0.9	-2.5	1.36
Washington	236	0.8	240	0.8	4.3 *	1.12
Wisconsin	224	1.4	222	2.3	-2.1	2.73
2005 and 2007 state assessment reported results are not comparable						
Connecticut	221	0.8	220	0.7	-0.8	1.12
Hawaii	247	1.0	238	0.5	-8.9 *	1.13
Idaho	207	2.2	217	0.9	10.2 *	2.34
Kansas	218	1.6	219	1.3	0.8	2.02
Maine	249	0.8	236	0.8	-12.8 *	1.13
Michigan	222	1.6	204	1.6	-18.3 *	2.23
Missouri	242	1.0	245	0.8	2.8 *	1.28
Montana	220	0.6	234	1.0	13.4 *	1.16
New York	207	1.4	219	0.8	12.0 *	1.60
North Carolina	203	0.9	231	0.6	28.4 *	1.12
Ohio	233	0.9	225	1.3	-8.1 *	1.57
Oklahoma	218	0.8	213	1.5	-5.1 *	1.68
West Virginia	215	1.1	217	1.3	2.2	1.66
Wyoming	251	0.8	216	0.6	-34.7 *	0.98

* Difference is statistically significant at $p < .05$.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table 18. Difference between the estimated NAEP scale equivalents of state grade 8 mathematics proficient standards and their standard error, by state: 2005 and 2007

State/jurisdiction	2005 NAEP scale equivalent	2005 Standard error	2007 NAEP scale equivalent	2007 Standard error	Difference 2007-2005	Standard error
2005 and 2007 state assessment reported results are comparable						
Alaska	268	1.2	265	1.2	-3.0	1.70
Arizona	265	1.0	268	1.1	2.7	1.46
Arkansas	288	0.7	277	1.3	-11.0 *	1.54
Colorado	258	1.7	259	1.3	1.2	2.19
Florida	269	1.1	266	0.9	-3.0 *	1.37
Georgia	255	0.9	243	1.7	-11.7 *	1.89
Indiana	266	0.9	266	1.6	0.7	1.80
Iowa	262	1.3	264	1.5	2.0	2.00
Louisiana	264	0.8	267	1.2	2.7	1.50
Maryland	276	1.2	278	1.5	1.9	1.94
Mississippi	262	1.4	262	0.9	0.5	1.62
Nevada	271	1.5	267	1.2	-3.8 *	1.90
New Jersey	273	1.1	272	0.8	-0.9	1.37
New Mexico	287	1.6	285	0.9	-1.3	1.79
North Dakota	277	0.9	279	0.8	2.1	1.22
Ohio	274	1.2	265	1.2	-9.2 *	1.64
Pennsylvania	272	0.6	271	1.0	-0.7	1.20
South Carolina	305	0.9	312	1.4	6.8 *	1.63
Tennessee	230	1.3	234	2.2	4.3	2.51
Texas	272	0.6	268	1.0	-4.2 *	1.21
Wisconsin	263	1.1	262	1.7	-1.5	2.00
2005 and 2007 state assessment reported results are not comparable						
Connecticut	257	1.6	252	2.0	-4.6	2.56
Delaware	275	0.9	272	0.9	-3.2 *	1.30
Hawaii	296	1.2	294	0.8	-2.1	1.39
Idaho	266	1.8	265	1.6	-0.8	2.43
Illinois	276	0.9	251	0.8	-25.1 *	1.25
Kentucky	285	1.1	279	0.7	-6.2 *	1.34
Maine	300	1.2	286	0.9	-13.9 *	1.47
Massachusetts	301	1.1	302	1.1	1.6	1.52
Michigan	269	1.3	260	1.5	-8.4 *	1.98
Missouri	311	1.3	289	1.2	-22.2 *	1.77
Montana	271	1.1	281	1.7	10.4 *	2.02
New York	275	0.8	273	1.1	-2.5	1.40
North Carolina	247	1.5	270	1.3	22.7 *	1.92
Oklahoma	258	0.7	249	1.1	-8.9 *	1.33
Oregon	269	1.2	262	1.2	-6.9 *	1.64
Virginia	253	1.0	259	1.6	6.1 *	1.84
West Virginia	253	0.9	253	1.0	0.4	1.36
Wyoming	293	1.0	279	0.8	-13.4 *	1.30

* Difference is statistically significant at $p < .05$.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

5 CORROBORATING STATE ASSESSMENT MEASURES OF ACHIEVEMENT CHANGE WITH NAEP

In this section, we compare the change from 2005 to 2007 in the percentage of students meeting the state's standard and the change from 2005 to 2007 in the percentage of students meeting the NAEP scale equivalent of the same state's standard. 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. Therefore, to compare NAEP and state changes in achievement from 2005 to 2007, the percentage of students reported to be meeting the state standard in 2007 is compared with the percentage of the NAEP students in 2007 that are above the NAEP scale equivalent of the state standard in 2005.

Described in Section 2, the statistic D is defined as the discrepancy between the change from 2005 to 2007 in the percentage meeting the state standard on the state test and the change in the same percentage when measured by NAEP.[27] If the statistical test indicates that D is different from zero, students gained more between 2005 and 2007 on either the NAEP skill domain or on the state-specific skill domain, depending on whether D is positive or negative. When D is greater than zero, the change from 2005 to 2007 on the state assessment is more positive (or less negative) than the change from 2005 to 2007 on NAEP. This could happen in two ways. If the percentage of students meeting the standard on the state test increased, the comparison with NAEP would show a smaller increase in NAEP's percentage (or even a decrease). If a smaller percentage of students met the standard on the state test, the comparison with NAEP would show a larger loss on NAEP.

When D is less than zero, the change on the state assessment is less positive (or more negative) than the change on NAEP. This could also happen in two ways. If more students met the standard on the state test over these 2 years, the comparison with NAEP would show that even more students gained on NAEP than on the state test. If fewer students met the standard on the state test over this period, the comparison with NAEP would show either a smaller loss or a gain in student achievement. A focus on state-specific content during instruction might lead to a positive value for D, whereas a focus on NAEP content might lead to a negative value for D.

---

[27] In Section 2, equation 5 defined D. Rearranging the terms in the equation, D can be rewritten as $(D_S - D_N)$, where D_S is the change from 2005 to 2007 in achievement measured by the state test, and D_N is the change from 2005 to 2007 in achievement measured by the mapping. When $D > 0$, i.e., $D_S > D_N$, the change from 2005 to 2007 on the state assessment is more positive (or less negative) than the change from 2005 to 2007 on NAEP. For $D < 0$, that is $D_S < D_N$, the change on the state assessment is less positive (or more negative) than the change on NAEP.

To use Wisconsin reading grade 4 as an example from table 19,

$$D_S = 79.5 - 82.8 = -3.3$$
$$D_N = 83.3 - 82.8 = .5$$
$$D = D_S - D_N = -3.8$$

For Wisconsin reading grade 8,

$$D_S = 82.7 - 85.8 = -3.1$$
$$D_N = 84.8 - 85.8 = -1.0$$
$$D = D_S - D_N = -2.1$$

In both situations, the changes on the state assessment are less positive (or more negative) than the changes on NAEP.

If either NAEP or a state test has substantively changed between the two years, then comparisons of achievement changes identified by the two tests are not warranted. In the years from 2005 to 2007, many states changed their state assessments as shown in the tables in appendix B, and finding values of *D* significantly different from zero in those cases is to be expected.

Tables 19 through 22 display comparisons limited to the states that reported no significant changes in their own assessments between 2005 and 2007 that are large enough to affect the direct comparability of the 2005 and the 2007 reported results.[28]

Table 19 shows that of the 22 states with comparable test results in grade 4 reading, 11 showed no statistically significant difference between NAEP and state assessment changes in achievement between 2005 and 2007 (Alaska, California, Colorado, Indiana, Iowa, Maryland, Massachusetts, Mississippi, New Mexico, North Carolina, and Ohio), 5 showed changes that are more positive than the changes measured by NAEP (Arkansas, Louisiana, North Dakota, South Carolina, and Texas), and 6 states showed changes that are less positive than those measured by NAEP (Alabama, Florida, New Jersey, Tennessee, Washington, and Wisconsin).

Table 19. NAEP and state assessment percentages meeting the state grade 4 reading proficient standard in 2007 based on 2005 standards, by state

State/jurisdiction	State percent at the standard in 2005[1]	NAEP percent at the 2005 standard in 2007	State percent at the standard in 2007	Difference D	Standard error of D
Alabama	82.4	88.8	85.3	-3.5 *	0.86
Alaska	79.2	81.0	80.4	-0.6	0.77
Arkansas	53.5	53.2	57.9	4.6 *	1.05
California	47.8	51.0	50.7	-0.3	0.73
Colorado	86.0	86.0	85.7	-0.3	0.78
Florida	70.8	76.2	69.5	-6.7 *	0.75
Indiana	72.3	76.6	76.1	-0.4	0.97
Iowa	77.3	82.9	81.5	-1.3	1.29
Louisiana	65.4	62.7	67.1	4.4 *	1.63
Maryland	82.0	86.4	86.9	0.5	1.00
Massachusetts	48.3	53.4	56.3	2.8	1.46
Mississippi	88.1	91.2	90.1	-1.2	0.65
New Jersey	81.0	88.2	81.7	-6.5 *	0.92
New Mexico	50.3	57.6	55.7	-1.9	1.31
North Carolina	82.4	84.5	85.0	0.5	0.74
North Dakota	76.5	79.6	81.8	2.2 *	1.12
Ohio	76.6	81.4	81.6	0.2	1.13
South Carolina	34.7	36.2	42.4	6.2 *	1.21
Tennessee	87.9	89.8	87.6	-2.2 *	0.79
Texas	80.6	81.6	83.5	1.9 *	0.81
Washington	79.6	79.4	75.1	-4.3 *	1.11
Wisconsin	82.8	83.3	79.5	-3.8 *	0.95

* Difference is statistically significant at $p < .05$.

[1] This matches the NAEP percentage meeting 2005 standard in 2005, by definition.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

[28] Appendix C presents the results for all states with available data.

Table 20. NAEP and state assessment percentages meeting the state grade 8 reading proficient standard in 2007 based on 2005 standards, by state

State/jurisdiction	State percent at the standard in 2005[1]	NAEP percent at the 2005 standard in 2007	State percent at the standard in 2007	Difference D	Standard error of D
Alabama	69.3	69.1	71.4	2.3 *	1.07
Alaska	81.8	81.3	79.3	-2.0 *	0.93
Arizona	63.2	63.6	62.8	-0.9	1.06
Arkansas	57.6	58.0	63.8	5.8 *	1.22
California	39.2	41.1	41.9	0.8	0.82
Colorado	85.9	87.9	87.0	-0.9	0.87
Florida	43.5	46.9	50.7	3.7 *	0.78
Illinois	72.5	73.0	80.8	7.8 *	1.22
Indiana	66.3	70.0	68.5	-1.5	0.96
Iowa	72.3	74.1	72.8	-1.3	1.17
Louisiana	54.0	55.0	60.5	5.5 *	1.40
Maryland	67.7	74.4	69.3	-5.1 *	1.17
Mississippi	57.3	56.2	50.5	-5.7 *	1.14
Nevada	52.7	52.1	57.6	5.5 *	0.98
New Jersey	73.8	75.7	74.0	-1.7	1.13
New Mexico	51.9	53.6	56.0	2.4 *	1.17
North Carolina	87.6	88.1	87.9	-0.2	0.76
North Dakota	72.2	71.4	76.5	5.2 *	1.46
Ohio	80.1	81.4	82.2	0.8	0.89
Pennsylvania	64.3	65.5	77.2	11.7 *	1.31
South Carolina	30.3	30.0	24.7	-5.3 *	1.19
Tennessee	87.4	88.0	92.5	4.4 *	0.81
Texas	83.4	86.5	87.8	1.3	0.68
Wisconsin	85.8	84.8	82.7	-2.1 *	0.90

* Difference is statistically significant at $p < .05$.

[1] This matches the NAEP percentage meeting 2005 standard in 2005, by definition.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

For grade 8 reading, table 20 shows that 9 of the 24 states with comparable assessments did not show statistically significant differences between NAEP and state assessment changes in achievement between 2005 and 2007 (Arizona, California, Colorado, Indiana, Iowa, New Jersey, North Carolina, Ohio, and Texas), 10 states showed changes that are more positive than the changes measured by NAEP (Alabama, Arkansas, Florida, Illinois, Louisiana, Nevada, New Mexico, North Dakota, Pennsylvania, and Tennessee), and 5 showed changes that are less positive than those measured by NAEP (Alaska, Maryland, Mississippi, South Carolina, and Wisconsin).

Table 21. NAEP and state assessment percentages meeting the state grade 4 mathematics proficient standard in 2007 based on 2005 standards, by state

State/jurisdiction	State percent at the standard in 2005[1]	NAEP percent at the 2005 standard in 2007	State percent at the standard in 2007	Difference D	Standard error of D
Alabama	74.0	77.3	78.7	1.4	1.29
Alaska	70.7	71.3	76.7	5.3 *	1.20
Arkansas	52.9	55.3	64.8	9.5 *	1.09
California	51.4	51.2	57.3	6.1 *	0.76
Colorado	89.7	90.1	90.1	0.0	0.67
Florida	63.1	68.9	69.7	0.8	1.09
Georgia	74.5	77.4	78.9	1.5	0.95
Indiana	72.3	79.9	77.0	-2.9 *	1.13
Iowa	79.5	82.6	82.2	-0.4	1.04
Louisiana	62.6	61.7	61.3	-0.4	1.48
Maryland	78.1	79.5	86.3	6.9 *	1.28
Massachusetts	38.5	47.0	48.6	1.5	1.78
Mississippi	78.8	79.3	81.0	1.8	1.06
New Jersey	80.7	84.5	85.3	0.8	0.98
New Mexico	38.6	46.7	46.1	-0.5	1.07
North Dakota	80.1	82.3	80.4	-1.9 *	0.85
South Carolina	38.9	39.8	41.7	1.9	1.07
Tennessee	86.8	88.3	89.2	0.9	0.82
Texas	81.7	82.5	84.9	2.3 *	0.88
Washington	60.5	62.9	56.9	-5.9 *	1.14
Wisconsin	74.1	75.6	76.1	0.5	1.33

* Difference is statistically significant at $p < .05$.

[1] This matches the NAEP percentage meeting 2005 standard in 2005, by definition.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table 21 shows that of the 21 states with comparable test results in grade 4 mathematics, 13 showed no statistically significant difference between NAEP and state assessment changes in achievement between 2005 and 2007 (Alabama, Colorado, Florida, Georgia, Iowa, Louisiana, Massachusetts, Mississippi, New Jersey, New Mexico, South Carolina, Tennessee, and Wisconsin), 5 states showed changes that are more positive than the changes measured by NAEP (Alaska, Arkansas, California, Maryland, and Texas), and 3 showed changes that are less positive than those measured by NAEP (Indiana, North Dakota, and Washington).

Table 22. NAEP and state assessment percentages meeting the state grade 8 mathematics proficient standard in 2007 based on 2005 standards, by state

State/jurisdiction	State percent at the standard in 2005[1]	NAEP percent at the 2005 standard in 2007	State percent at the standard in 2007	Difference D	Standard error of D
Alaska	65.1	67.2	70.0	2.9 *	1.12
Arizona	60.5	63.1	60.2	-2.9 *	1.12
Arkansas	33.7	36.3	48.4	12.1 *	1.34
Colorado	74.1	78.7	77.6	-1.0	0.94
Florida	58.2	60.8	64.1	3.4 *	0.92
Georgia	68.7	71.9	82.5	10.6 *	1.35
Indiana	70.2	72.2	71.5	-0.7	1.23
Iowa	75.6	77.6	75.6	-2.0	1.09
Louisiana	56.3	62.3	58.7	-3.6 *	1.37
Maryland	53.0	60.4	58.4	-2.0	1.22
Mississippi	52.5	54.4	53.6	-0.8	1.13
Nevada	51.1	49.6	53.8	4.1 *	0.89
New Jersey	63.9	67.8	68.5	0.7	1.10
New Mexico	23.6	28.5	29.7	1.2	0.92
North Dakota	65.5	70.6	68.0	-2.6 *	1.28
Ohio	62.7	65.0	74.0	9.0 *	1.22
Pennsylvania	62.4	68.9	69.7	0.8	1.16
South Carolina	23.8	25.7	19.9	-5.7 *	1.18
Tennessee	87.8	90.8	88.5	-2.3 *	0.91
Texas	60.9	67.4	71.9	4.5 *	0.99
Wisconsin	74.9	75.4	74.4	-1.0	1.16

* Difference is statistically significant at $p < .05$.

[1] This matches the NAEP percentage meeting 2005 standard in 2005, by definition.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

For grade 8 mathematics, table 22 shows that 9 out of 21 states with comparable assessments showed no statistically significant difference between NAEP and state assessment measures of changes in achievement between 2005 and 2007 (Colorado, Indiana, Iowa, Maryland, Mississippi, New Jersey, New Mexico, Pennsylvania, and Wisconsin), 7 states with changes that are more positive than the changes measured by NAEP (Alaska, Arkansas, Florida, Georgia, Nevada, Ohio, Texas), and 5 states with changes that are less positive than those measured by NAEP (Arizona, Louisiana, North Dakota, South Carolina, and Tennessee).

Tables 23 through 25 summarize the results by listing where NAEP and state assessment do and do not agree. Table 23 lists the states that show changes in achievement in their own test that are corroborated by NAEP results, 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. Table 24 lists the states showing more positive changes in student achievement from 2005 to 2007 than NAEP, and table 25 lists the states with less positive changes than NAEP from 2005 to 2007.

It is important to understand the reasons for the discrepancies. Because of the complexity of testing, in most cases, the source of the discrepancy (or drift) is likely to be some change in testing, such as in accommodation, exclusions, time of testing, or scaling methods. Even when these sources are ruled out, differences in the domains covered by the two tests can lead to discrepancies in achievement changes. While it is beyond the scope of this report to undertake such analyses, it may be valuable for the states where such differences exist to do so.

Table 23. States showing changes in student achievement from 2005 to 2007 in their own tests that are corroborated by NAEP results in the same period, by subject and grade

Reading		Mathematics	
Grade 4	**Grade 8**	**Grade 4**	**Grade 8**
Alaska	Arizona	Alabama	Colorado
California	California	Colorado	Indiana
Colorado	Colorado	Florida	Iowa
Indiana	Indiana	Georgia	Maryland
Iowa	Iowa	Iowa	Mississippi
Maryland	New Jersey	Louisiana	New Jersey
Massachusetts	North Carolina	Massachusetts	New Mexico
Mississippi	Ohio	Mississippi	Pennsylvania
New Mexico	Texas	New Jersey	Wisconsin
North Carolina		New Mexico	
Ohio		South Carolina	
		Tennessee	
		Wisconsin	

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading and Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table 24. States showing changes in student achievement from 2005 to 2007 in their own tests that are statistically significantly more positive than NAEP's, by subject and grade

Reading		Mathematics	
Grade 4	**Grade 8**	**Grade 4**	**Grade 8**
Arkansas	Alabama	Alaska	Alaska
Louisiana	Arkansas	Arkansas	Arkansas
North Dakota	Florida	California	Florida
South Carolina	Illinois	Maryland	Georgia
Texas	Louisiana	Texas	Nevada
	Nevada		Ohio
	New Mexico		Texas
	North Dakota		
	Pennsylvania		
	Tennessee		

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading and Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table 25. States showing changes in student achievement from 2005 to 2007 in their own tests that are statistically significantly less positive than NAEP's, by subject and grade

Reading		Mathematics	
Grade 4	**Grade 8**	**Grade 4**	**Grade 8**
Alabama	Alaska	Indiana	Arizona
Florida	Maryland	North Dakota	Louisiana
New Jersey	Mississippi	Washington	North Dakota
Tennessee	South Carolina		South Carolina
Washington	Wisconsin		Tennessee
Wisconsin			

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading and Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

6 CONCLUSIONS

The mapping results described in this study have made it possible to compare state reading and mathematics proficiency standards across states, using grade 4 and 8 NAEP reading and mathematics scales as common yardsticks. The findings have also made it possible to evaluate consistency in the state standards over time and use NAEP to corroborate progress (or lack thereof) in the achievement assessed by states. Identifying a NAEP scale equivalent score for the state's standard was an essential step for the analyses conducted in this study. These analyses were based on school-level percentages of students meeting a state's standard on a state's own tests, which are systematically available for almost every state and could be compared with student performance on NAEP in the same schools.

The purpose of state-to-NAEP comparisons is to aid in the interpretation of state assessment results by providing a benchmark. Despite the limitations of state-to-NAEP comparisons, there is a need for reliable information that compares state standards to one another. What does it mean to say that a student is proficient in reading in grade 4 in Massachusetts? Would a fourth-grader who is proficient in reading in Wyoming also be proficient in Massachusetts? The analyses presented in this study provide a basis for answering such questions.

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; and in mathematics, only two states' NAEP scale equivalent were in the NAEP *Proficient* range (Massachusetts in grades 4 and 8, and South Carolina in grade 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.

Almost one-half of the states changed aspects of their assessment policies or the assessment itself between 2005 and 2007 in ways that prevented their reading or mathematics test results from being comparable across these two years. Either explicitly or implicitly, such states have 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. For states with both years of data, the mapping results showed that the NAEP equivalents representing state standards for proficiency were lower in 2007 in one-third to one-half of the states that made such changes (depending on subject and grade). A decrease in the stringency of the NAEP equivalent of state standards was more likely to occur for grade 8 than for grade 4.

In the remaining states in which no changes were made or the changes in assessment policies were minor enough that their test results remained comparable, it was possible to check the extent to which NAEP corroborates the changes in achievement measured in the states' assessments. In two-fifths to three-fifths of the states (depending on subject and grade), NAEP's measurements of student progress agreed with the progress measured by state assessments. In cases in which NAEP and the state disagreed on their measurement of student progress, the

findings could both be accurate, as the underlying domains of the two tests may not involve the same skills or the same skills in equal weights. Similarly, there may have been a methodological change between 2005 and 2007 in the state tests, in such areas as exclusions, time of administrations, or scaling.

In all 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 this study show that NAEP, as a common yardstick, is an essential benchmark for states in evaluating their standards.

REFERENCES

Braun, H.I., and Qian, J. (2007). *Mapping State Performance Standards Onto the NAEP Scale*. Princeton, NJ: Educational Testing Service.

Feuer, M.J., Holland, P.W., Green, B.F., Bertenthal, M.W., and Hemphill, F.C. (Eds). (1999). *Uncommon Measures: Equivalency and Linkage of Educational Tests* (Committee on Equivalency and Linkage of Educational Tests, National Research Council). Washington, DC: The National Academies Press.

Ho, A., and Haertel, E. (2007). *Apples to Apples: The Underlying Assumptions of State-NAEP Comparisons* (Research brief). Retrieved July 14, 2007, from http://www.ccsso.org/content/PDFs/Ho Haertel CCSSO Brief2 Final.pdf.

Koretz, D.M., Bertenthal, M.W., and Green, B.F. (Eds). (1999). *Embedding Questions: The Pursuit of a Common Measure in Uncommon Tests*, (Committee on Embedding Common Test Items in State and District Assessments, National Research Council). Washington, DC: The National Academies Press.

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

McLaughlin, D., and Bandeira de Mello, V. (2002). *Comparison of State Elementary School Mathematics Achievement Standards, Using NAEP 2000*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

McLaughlin, D., and Bandeira de Mello, V. (2003). *Comparing State Reading and Math Performance Standards Using NAEP*. Paper presented at the annual National Conference on Large-Scale Assessment, San Antonio.

McLaughlin, D., and Bandeira de Mello, V. (2006). *How to Compare NAEP and State Assessment Results*. Paper presented at the annual National Conference on Large-Scale Assessment. Retrieved December 2007, from http://www.naepreports.org/task 1.1/LSAC_20050618.ppt.

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.

APPENDIX A
NUMBER OF SCHOOLS IN THE NAEP SAMPLE AND THE PERCENTAGE OF SCHOOLS USED IN THE 2007 MAPPING

Sample sizes and percentages of the 2007 NAEP samples used in comparisons are shown in tables A-1 and A-2 for reading and mathematics, respectively. For each grade, the tables display the number of public schools selected for NAEP in each state, the percentage of these schools included in the analyses in this report, and the percentage of the student population represented by the comparison schools. The percentage of the population represented by NAEP can be less than 100 percent either because of failure to match schools in the two databases or because scores for the school are suppressed on the data source. In general, because the schools missing state assessment scores are generally small schools, the percentages of student populations represented by the school used in the comparisons are generally higher than the percentages of schools.

Table A-1. Number of NAEP schools, percentage of NAEP schools available for comparing state assessment results with NAEP results in grades 4 and 8 reading, and the percentage of the student population in these comparison schools, by state: 2007

State/jurisdiction	Grade 4			Grade 8		
	NAEP schools[1]	Percent of NAEP schools matched	Percent of population represented	NAEP schools[1]	Percent of NAEP schools matched	Percent of population represented
Alabama	110	99.1	99.1	120	100.0	100.0
Alaska	180	99.4	99.9	110	98.2	99.3
Arizona	120	100.0	100.0	130	97.7	99.2
Arkansas	120	96.6	97.6	120	91.1	94.5
California	320	97.8	99.0	310	97.1	99.0
Colorado	120	95.8	99.0	120	93.1	98.5
Connecticut	110	100.0	100.0	100	100.0	100.0
Delaware	100	100.0	100.0	50	97.8	100.0
District of Columbia	†	—	†	†	—	†
Florida	160	97.6	97.2	160	98.7	98.6
Georgia	160	98.7	96.4	120	97.5	95.6
Hawaii	120	99.1	99.1	70	100.0	100.0
Idaho	130	95.5	91.2	110	97.2	99.0
Illinois	180	98.9	99.6	200	98.0	99.3
Indiana	110	100.0	100.0	110	100.0	100.0
Iowa	140	97.8	96.7	130	97.0	96.8
Kansas	140	98.6	99.0	150	97.3	98.0
Kentucky	120	97.4	98.1	110	98.2	98.3
Louisiana	110	97.2	98.4	110	96.4	97.7
Maine	150	93.4	95.4	130	94.7	97.4
Maryland	110	98.2	98.4	110	99.1	97.3
Massachusetts	170	100.0	100.0	140	99.3	99.4
Michigan	120	99.2	98.7	120	96.7	98.3
Minnesota	130	100.0	100.0	140	97.9	98.4
Mississippi	120	97.4	97.1	110	97.4	97.9
Missouri	130	98.4	99.4	130	94.7	95.8
Montana	190	98.9	99.1	170	98.2	99.4
Nebraska	†	—	†	†	—	†
Nevada	110	94.5	94.0	70	93.2	93.3
New Hampshire	130	89.9	92.2	90	96.7	99.2
New Jersey	110	98.2	95.1	110	100.0	100.0
New Mexico	130	95.3	97.9	110	97.3	99.4
New York	150	99.3	99.8	160	98.1	98.5
North Carolina	170	97.6	96.5	150	99.3	99.8
North Dakota	210	80.5	93.3	190	70.3	90.0
Ohio	160	98.1	99.3	190	98.4	99.1
Oklahoma	140	98.5	98.8	150	96.6	96.8
Oregon	140	97.0	98.9	110	96.5	99.0
Pennsylvania	110	99.1	98.5	110	98.2	97.6
Rhode Island	110	100.0	100.0	60	100.0	100.0
South Carolina	110	97.2	98.7	110	97.2	98.5
South Dakota	190	98.4	98.0	140	99.3	99.7
Tennessee	120	100.0	100.0	120	99.2	99.1
Texas	300	98.6	97.9	220	96.4	97.4
Utah	†	—	†	†	—	†
Vermont	190	85.4	93.9	120	86.8	97.5
Virginia	110	97.4	97.1	110	99.1	98.9
Washington	130	99.2	100.0	130	100.0	100.0
West Virginia	150	92.5	89.7	120	91.5	91.1
Wisconsin	130	65.4	71.0	130	74.6	82.1
Wyoming	170	96.5	97.2	80	95.1	96.1

— State assessment data not available.

† Not applicable.

[1] Rounded to the nearest 10 for confidentiality.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table A-2. Number of NAEP schools, percentage of NAEP schools available for comparing state assessment results with NAEP results in grades 4 and 8 mathematics, and percentage of the student population in these comparison schools, by state: 2007

State/jurisdiction	Grade 4			Grade 8		
	NAEP schools[1]	Percent of NAEP schools matched	Percent of population represented	NAEP schools[1]	Percent of NAEP schools matched	Percent of population represented
Alabama	110	99.1	99.1	120	100.0	100.0
Alaska	180	100.0	100.0	110	98.2	99.3
Arizona	120	100.0	100.0	130	97.7	99.1
Arkansas	120	96.6	97.5	130	90.4	94.5
California	330	97.5	98.9	310	95.5	97.8
Colorado	120	95.8	99.1	120	93.1	98.6
Connecticut	110	100.0	100.0	100	100.0	100.0
Delaware	100	100.0	100.0	50	100.0	100.0
District of Columbia	†	—	†	†	—	†
Florida	160	97.6	97.2	160	98.7	98.6
Georgia	160	98.7	96.4	120	97.5	95.4
Hawaii	120	99.1	99.0	70	95.7	99.8
Idaho	130	95.5	91.8	100	98.1	99.1
Illinois	180	98.9	99.5	200	98.0	99.4
Indiana	110	100.0	100.0	110	100.0	100.0
Iowa	140	97.8	96.9	140	96.3	96.9
Kansas	140	98.6	99.0	150	97.3	98.0
Kentucky	120	97.4	98.2	110	98.2	98.7
Louisiana	110	97.2	98.4	110	96.4	97.7
Maine	150	93.4	95.6	130	94.7	97.6
Maryland	110	98.2	98.5	110	99.1	97.2
Massachusetts	170	100.0	100.0	130	99.3	99.2
Michigan	120	99.2	98.8	120	96.7	97.9
Minnesota	130	100.0	100.0	140	98.5	98.2
Mississippi	120	97.4	97.2	110	97.4	97.7
Missouri	130	98.4	99.6	130	94.7	96.2
Montana	190	98.9	99.3	170	98.2	99.4
Nebraska	†	—	†	†	—	†
Nevada	110	94.6	93.6	80	93.3	93.6
New Hampshire	130	89.9	92.1	90	97.8	99.2
New Jersey	110	98.2	95.0	110	100.0	100.0
New Mexico	130	93.8	97.5	110	97.3	99.6
New York	150	99.3	99.8	160	98.1	98.5
North Carolina	170	97.6	96.4	150	99.3	99.7
North Dakota	210	81.3	93.1	180	70.3	89.6
Ohio	160	98.1	99.4	190	98.9	98.8
Oklahoma	140	98.6	98.7	150	96.6	96.8
Oregon	140	97.0	98.9	110	96.5	99.2
Pennsylvania	110	99.1	98.6	110	98.2	97.5
Rhode Island	110	100.0	100.0	60	100.0	100.0
South Carolina	110	97.2	98.3	110	97.2	98.8
South Dakota	190	98.4	98.1	140	99.3	99.7
Tennessee	120	100.0	100.0	120	99.2	99.2
Texas	300	98.6	98.0	220	96.4	97.6
Utah	†	—	†	†	—	†
Vermont	190	85.0	93.7	120	86.8	97.0
Virginia	110	97.4	97.4	110	100.0	100.0
Washington	130	99.2	100.0	130	100.0	100.0
West Virginia	150	92.5	89.5	120	91.5	91.0
Wisconsin	130	65.4	70.7	130	74.6	82.6
Wyoming	170	97.6	97.2	80	96.3	97.1

— State assessment data not available.

† Not applicable.

[1] Rounded to the nearest 10 for confidentiality.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

APPENDIX B
CHANGES IN STATES' ASSESSMENTS BETWEEN 2005 AND 2007

Tables B-1 and B-2 summarize selected changes in states' assessments between the two NAEP administrations of 2005 and 2007. Their source is the *2007 Survey of State Assessment Program Characteristics*, a survey designed to provide contextual information to document general state assessment program information, mainly from the section covering changes that were made to state assessments between the 2004-05 and 2006-07 school years. States were instructed to indicate whether they had added grades, eliminated grades, changed cut scores, changed the time of year when the test was administered, changed the assessment items significantly, used an entirely new assessment, realigned the assessment to new content standards, changed the proficiency standards, changed the accommodation policy, changed the re-test policy, or changed test contractors. Additionally, states could indicate that there were *no significant changes* to the state assessment between 2004-05 and 2006-07 or, if applicable, to describe any changes in further detail.

States were also asked to indicate whether the following statement was true or false for grades 4 and 8 Reading/Language Arts and Mathematics:

> *The reported 2006–07 state assessment results for 4th- and 8th-grade Reading and Mathematics are directly comparable with the 2004–05 reported results.*

Finally, states were asked to indicate whether there were any policy or legislative changes in the administration of the Reading/Language Arts and Mathematics assessments or in the reporting of outcomes between 2004-05 and 2006-07 that would have an impact on the interpretation of school- or state-level results when comparing across years. Table B-3 summarizes these responses.

State profiles tabulating the survey results are available at http://nces.ed.gov/nationsreportcard/studies/statemapping.asp.

Table B-1. Selected changes to state reading assessments between the 2004–05 and the 2006–07 administrations, by state

State/jurisdiction	Added grades	Eliminated grades	Changed cut scores	Changed the time of administration	Changed assessment items	Entirely different assessment
Alabama					√	
Alaska	√					
Arizona					√	
Arkansas					√	
California					√	
Colorado						
Connecticut	√		√	√		
Delaware	√		√			
District of Columbia	√		√			√
Florida						
Georgia			√		√	
Hawaii	√		√		√	√
Idaho			√		√	√
Illinois	√				√	√
Indiana						
Iowa	√					
Kansas	√		√		√	√
Kentucky	√		√		√	
Louisiana						√
Maine	√		√	√	√	√
Maryland						
Massachusetts	√			√	√	
Michigan	√		√	√	√	
Minnesota	√					√
Mississippi						
Missouri	√		√			√
Montana	√		√			
Nebraska	√		√		√	
Nevada	√					
New Hampshire	√		√	√	√	√
New Jersey	√					
New Mexico						
New York	√		√			
North Carolina						
North Dakota						
Ohio						
Oklahoma	√					
Oregon	√		√			
Pennsylvania	√		√		√	
Rhode Island						
South Carolina						
South Dakota			√		√	
Tennessee						
Texas						
Utah		√			√	
Vermont	√					
Virginia	√		√			
Washington	√					
West Virginia						√
Wisconsin	√					
Wyoming	√		√	√	√	√

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2007 Survey of State Assessment Program Characteristics.

Table B-1. Selected changes to state reading assessments between the 2004–05 and the 2006–07 administrations, by state—Continued

State/jurisdiction	Realigned to new content standards	Changed proficiency standards	Changed accommodation policy	Changed re-test policy	Changed test contractors	No significant changes
Alabama						√
Alaska						
Arizona						
Arkansas			√			
California						
Colorado						√
Connecticut	√				√	
Delaware		√				
District of Columbia	√				√	
Florida					√	
Georgia	√	√			√	
Hawaii	√	√	√		√	
Idaho	√	√			√	
Illinois					√	
Indiana						√
Iowa						
Kansas	√	√				
Kentucky	√	√			√	
Louisiana						
Maine	√	√	√		√	
Maryland						√
Massachusetts						
Michigan	√	√			√	
Minnesota	√	√			√	
Mississippi						√
Missouri		√	√			
Montana						
Nebraska						
Nevada						
New Hampshire	√	√	√			
New Jersey						
New Mexico						√
New York	√	√				
North Carolina						√
North Dakota						√
Ohio						√
Oklahoma						
Oregon		√				
Pennsylvania	√	√			√	
Rhode Island						√
South Carolina						√
South Dakota						
Tennessee						√
Texas						√
Utah	√		√			
Vermont						
Virginia	√	√			√	
Washington						
West Virginia						
Wisconsin						
Wyoming		√	√	√	√	

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2007 Survey of State Assessment Program Characteristics.

Table B-2. Selected changes to state mathematics assessments between the 2004–05 and the 2006–07 administrations, by state

State/jurisdiction	Added grades	Eliminated grades	Changed cut scores	Changed the time of administration	Changed assessment items	Entirely different assessment
Alabama					√	
Alaska	√					
Arizona					√	
Arkansas					√	
California					√	
Colorado						
Connecticut	√		√	√		
Delaware	√		√			
District of Columbia	√		√			√
Florida						
Georgia			√		√	
Hawaii	√		√		√	√
Idaho			√		√	√
Illinois	√		√		√	√
Indiana						
Iowa	√					
Kansas	√		√		√	√
Kentucky	√		√		√	
Louisiana						√
Maine	√		√	√	√	√
Maryland						
Massachusetts	√				√	
Michigan	√		√	√	√	
Minnesota	√				√	√
Mississippi						
Missouri	√		√			√
Montana	√		√			
Nebraska	√		√		√	
Nevada	√					
New Hampshire	√		√	√	√	√
New Jersey	√					
New Mexico						
New York	√		√			
North Carolina			√		√	√
North Dakota						
Ohio						√
Oklahoma	√					
Oregon	√		√			
Pennsylvania	√		√		√	
Rhode Island						
South Carolina						
South Dakota			√		√	
Tennessee						
Texas						
Utah		√			√	
Vermont	√					
Virginia	√		√			
Washington	√					
West Virginia						√
Wisconsin	√					
Wyoming	√		√	√	√	√

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2007 Survey of State Assessment Program Characteristics.

Table B-2. Selected changes to state mathematics assessments between the 2004–05 and the 2006–07 administrations, by state—Continued

State/jurisdiction	Realigned to new content standards	Changed proficiency standards	Changed accommodation policy	Changed re-test policy	Changed test contractors	No significant changes
Alabama						√
Alaska						
Arizona						
Arkansas			√			
California						
Colorado						√
Connecticut	√				√	
Delaware		√				
District of Columbia	√				√	
Florida					√	
Georgia	√	√			√	
Hawaii	√	√			√	
Idaho	√	√			√	
Illinois					√	
Indiana						√
Iowa						
Kansas	√	√				
Kentucky	√	√			√	
Louisiana						
Maine	√	√	√		√	
Maryland						√
Massachusetts						
Michigan	√	√			√	
Minnesota	√	√			√	
Mississippi						√
Missouri		√				
Montana						
Nebraska						
Nevada						
New Hampshire	√	√	√			
New Jersey						
New Mexico						√
New York	√	√				
North Carolina	√	√				
North Dakota						√
Ohio						
Oklahoma						
Oregon		√				
Pennsylvania	√	√			√	
Rhode Island						√
South Carolina						√
South Dakota	√					
Tennessee						√
Texas						√
Utah			√		√	
Vermont						
Virginia	√	√			√	
Washington						
West Virginia						
Wisconsin						
Wyoming		√	√	√	√	

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2007 Survey of State Assessment Program Characteristics.

Table B-3. Comparability of the 2007 state assessment results in reading and mathematics at grades 4 and 8 with the 2005 reported results, by state

State/jurisdiction	Reading		Mathematics	
	Grade 4	Grade 8	Grade 4	Grade 8
Alabama	Yes	Yes	Yes	Yes
Alaska	Yes	Yes	Yes	Yes
Arizona	Yes	Yes	Yes	Yes
Arkansas	Yes	Yes	Yes	Yes
California	Yes	Yes	Yes	Yes
Colorado	Yes	Yes	Yes	Yes
Connecticut	No	No	No	No
Delaware	No	No	No	No
District of Columbia	No	No	No	No
Florida	Yes	Yes	Yes	Yes
Georgia	No	No	Yes	Yes
Hawaii	No	No	No	No
Idaho	No	No	No	No
Illinois	No	Yes	No	No
Indiana	Yes	Yes	Yes	Yes
Iowa	Yes	Yes	Yes	Yes
Kansas	No	No	No	No
Kentucky	No	No	No	No
Louisiana	Yes	Yes	Yes	Yes
Maine	No	No	No	No
Maryland	Yes	Yes	Yes	Yes
Massachusetts	Yes	No	Yes	No
Michigan	No	No	No	No
Minnesota	No	No	No	No
Mississippi	Yes	Yes	Yes	Yes
Missouri	No	No	No	No
Montana	No	No	No	No
Nebraska	No	No	No	No
Nevada	No	Yes	No	Yes
New Hampshire	No	No	No	No
New Jersey	Yes	Yes	Yes	Yes
New Mexico	Yes	Yes	Yes	Yes
New York	No	No	No	No
North Carolina	Yes	Yes	No	No
North Dakota	Yes	Yes	Yes	Yes
Ohio	Yes	Yes	No	Yes
Oklahoma	No	No	No	No
Oregon	No	No	No	No
Pennsylvania	No	Yes	No	Yes
Rhode Island	Yes	Yes	Yes	Yes
South Carolina	Yes	Yes	Yes	Yes
South Dakota	Yes	Yes	Yes	Yes
Tennessee	Yes	Yes	Yes	Yes
Texas	Yes	Yes	Yes	Yes
Utah	Yes	No	Yes	Yes
Vermont	Yes	Yes	Yes	Yes
Virginia	No	No	No	No
Washington	Yes	No	Yes	No
West Virginia	No	No	No	No
Wisconsin	Yes	Yes	Yes	Yes
Wyoming	No	No	No	No

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2007 Survey of State Assessment Program Characteristics.

APPENDIX C
SUPPLEMENTARY TABLES

Tables C-1 through C-4 are equivalent to tables 19 through 22, respectively, but with the additional results for states with changes in their state assessments between 2005 and 2007.

Tables C-5 and C-6, for reading and mathematics, respectively, list the number of states according to the statistical significance of the difference *D* and by whether changes in the state's own assessments between 2005 and 2007 were deemed to affect the direct comparability of the 2005 and the 2007 reported results.

For the states with both years of data that are comparable according to state assessment staff, tables C-7 through C-10 list selected changes to state assessments between 2005 and 2007, by whether reports of achievement changes from 2005 to 2007 in the state test and NAEP agree.

Table C-1. NAEP and state assessment percentages meeting the state grade 4 reading proficient standards in 2007 based on 2005 standards, by state

State/jurisdiction	State percent at the standard in 2005[1]	NAEP percent at the 2005 standard in 2007	State percent at the standard in 2007	Difference D	Standard error of D
2005 and 2007 state assessment reported results are comparable					
Alabama	82.4	88.8	85.3	-3.5 *	0.86
Alaska	79.2	81.0	80.4	-0.6	0.77
Arkansas	53.5	53.2	57.9	4.6 *	1.05
California	47.8	51.0	50.7	-0.3	0.73
Colorado	86.0	86.0	85.7	-0.3	0.78
Florida	70.8	76.2	69.5	-6.7 *	0.75
Indiana	72.3	76.6	76.1	-0.4	0.97
Iowa	77.3	82.9	81.5	-1.3	1.29
Louisiana	65.4	62.7	67.1	4.4 *	1.63
Maryland	82.0	86.4	86.9	0.5	1.00
Massachusetts	48.3	53.4	56.3	2.8	1.46
Mississippi	88.1	91.2	90.1	-1.2	0.65
New Jersey	81.0	88.2	81.7	-6.5 *	0.92
New Mexico	50.3	57.6	55.7	-1.9	1.31
North Carolina	82.4	84.5	85.0	0.5	0.74
North Dakota	76.5	79.6	81.8	2.2 *	1.12
Ohio	76.6	81.4	81.6	0.2	1.13
South Carolina	34.7	36.2	42.4	6.2 *	1.21
Tennessee	87.9	89.8	87.6	-2.2 *	0.79
Texas	80.6	81.6	83.5	1.9 *	0.81
Washington	79.6	79.4	75.1	-4.3 *	1.11
Wisconsin	82.8	83.3	79.5	-3.8 *	0.95
2005 and 2007 state assessment reported results are not comparable					
Connecticut	66.4	69.1	68.4	-0.6	1.06
Georgia	86.5	91.2	85.2	-6.0 *	0.68
Hawaii	56.4	62.3	54.6	-7.6 *	1.14
Idaho	86.9	88.0	80.4	-7.6 *	0.77
Kentucky	67.0	70.1	71.7	1.6	1.29
Maine	52.8	54.4	67.1	12.7 *	1.13
Michigan	83.4	85.3	87.4	2.0 *	0.98
Montana	80.6	83.9	79.9	-4.0 *	0.90
New York	70.5	70.8	68.4	-2.4 *	1.14
Oklahoma	82.2	86.3	90.9	4.6 *	0.97
West Virginia	80.4	80.8	83.2	2.4 *	0.97
Wyoming	46.9	49.7	76.9	27.1 *	0.91

* Difference is statistically significant at $p < .05$.

[1] This matches the NAEP percentage meeting 2005 standard in 2005, by definition.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table C-2. NAEP and state assessment percentages meeting the state grade 8 reading proficient standards in 2007 based on 2005 standards, by state

State/jurisdiction	State percent at the standard in 2005[1]	NAEP percent at the 2005 standard in 2007	State percent at the standard in 2007	Difference D	Standard error of D
2005 and 2007 state assessment reported results are comparable					
Alabama	69.3	69.1	71.4	2.3 *	1.07
Alaska	81.8	81.3	79.3	-2.0 *	0.93
Arizona	63.2	63.6	62.8	-0.9	1.06
Arkansas	57.6	58.0	63.8	5.8 *	1.22
California	39.2	41.1	41.9	0.8	0.82
Colorado	85.9	87.9	87.0	-0.9	0.87
Florida	43.5	46.9	50.7	3.7 *	0.78
Illinois	72.5	73.0	80.8	7.8 *	1.22
Indiana	66.3	70.0	68.5	-1.5	0.96
Iowa	72.3	74.1	72.8	-1.3	1.17
Louisiana	54.0	55.0	60.5	5.5 *	1.40
Maryland	67.7	74.4	69.3	-5.1 *	1.17
Mississippi	57.3	56.2	50.5	-5.7 *	1.14
Nevada	52.7	52.1	57.6	5.5 *	0.98
New Jersey	73.8	75.7	74.0	-1.7	1.13
New Mexico	51.9	53.6	56.0	2.4 *	1.17
North Carolina	87.6	88.1	87.9	-0.2	0.76
North Dakota	72.2	71.4	76.5	5.2 *	1.46
Ohio	80.1	81.4	82.2	0.8	0.89
Pennsylvania	64.3	65.5	77.2	11.7 *	1.31
South Carolina	30.3	30.0	24.7	-5.3 *	1.19
Tennessee	87.4	88.0	92.5	4.4 *	0.81
Texas	83.4	86.5	87.8	1.3	0.68
Wisconsin	85.8	84.8	82.7	-2.1 *	0.90
2005 and 2007 state assessment reported results are not comparable					
Connecticut	76.7	77.4	75.3	-2.1 *	0.99
Delaware	80.5	78.2	80.0	1.8 *	0.78
Georgia	82.6	85.5	89.8	4.3 *	0.94
Hawaii	37.3	41.0	60.4	19.4 *	1.09
Idaho	82.0	84.6	86.1	1.5 *	0.73
Kansas	78.0	81.2	82.1	0.9	0.93
Maine	44.1	44.7	64.4	19.7 *	1.73
Montana	72.2	76.7	79.2	2.6 *	1.01
New York	48.6	47.8	57.4	9.6 *	1.08
Oklahoma	71.2	71.3	81.4	10.1 *	1.09
Oregon	63.7	67.6	70.5	2.9 *	1.19
Virginia	78.2	78.4	82.0	3.6 *	1.22
West Virginia	79.8	80.4	80.4	0.0	1.28
Wyoming	39.0	37.7	72.8	35.1 *	1.11

* Difference is statistically significant at $p < .05$.

[1] This matches the NAEP percentage meeting 2005 standard in 2005, by definition.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table C-3. NAEP and state assessment percentages meeting the state grade 4 mathematics proficient standards in 2007 based on 2005 standards, by state

State/jurisdiction	State percent at the standard in 2005[1]	NAEP percent at the 2005 standard in 2007	State percent at the standard in 2007	Difference D	Standard error of D
2005 and 2007 state assessment reported results are comparable					
Alabama	74.0	77.3	78.7	1.4	1.29
Alaska	70.7	71.3	76.7	5.3 *	1.20
Arkansas	52.9	55.3	64.8	9.5 *	1.09
California	51.4	51.2	57.3	6.1 *	0.76
Colorado	89.7	90.1	90.1	0.0	0.67
Florida	63.1	68.9	69.7	0.8	1.09
Georgia	74.5	77.4	78.9	1.5	0.95
Indiana	72.3	79.9	77.0	-2.9 *	1.13
Iowa	79.5	82.6	82.2	-0.4	1.04
Louisiana	62.6	61.7	61.3	-0.4	1.48
Maryland	78.1	79.5	86.3	6.9 *	1.28
Massachusetts	38.5	47.0	48.6	1.5	1.78
Mississippi	78.8	79.3	81.0	1.8	1.06
New Jersey	80.7	84.5	85.3	0.8	0.98
New Mexico	38.6	46.7	46.1	-0.5	1.07
North Dakota	80.1	82.3	80.4	-1.9 *	0.85
South Carolina	38.9	39.8	41.7	1.9	1.07
Tennessee	86.8	88.3	89.2	0.9	0.82
Texas	81.7	82.5	84.9	2.3 *	0.88
Washington	60.5	62.9	56.9	-5.9 *	1.14
Wisconsin	74.1	75.6	76.1	0.5	1.33
2005 and 2007 state assessment reported results are not comparable					
Connecticut	78.2	78.5	79.2	0.6	0.89
Hawaii	29.6	36.2	48.9	12.7 *	0.92
Idaho	90.6	89.2	82.2	-7.0 *	0.76
Kansas	85.3	86.8	86.2	-0.6	1.08
Maine	39.8	42.3	61.2	19.0 *	1.18
Michigan	73.0	72.6	87.3	14.7 *	1.24
Missouri	40.9	48.2	44.3	-3.8 *	1.49
Montana	79.8	82.3	67.8	-14.5 *	1.07
New York	86.6	89.1	80.9	-8.2 *	0.72
North Carolina	91.5	92.0	66.5	-25.4 *	0.84
Ohio	65.2	69.2	78.2	9.1 *	1.44
Oklahoma	73.7	78.1	83.2	5.1 *	1.33
West Virginia	74.7	80.8	78.4	-2.4 *	1.14
Wyoming	39.2	41.0	87.0	46.0 *	0.92

* Difference is statistically significant at $p < .05$.

[1] This matches the NAEP percentage meeting 2005 standard in 2005, by definition.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table C-4. NAEP and state assessment percentages meeting the state grade 8 mathematics proficient standards in 2007 based on 2005 standards, by state

State/jurisdiction	State percent at the standard in 2005[1]	NAEP percent at the 2005 standard in 2007	State percent at the standard in 2007	Difference D	Standard error of D
2005 and 2007 state assessment reported results are comparable					
Alaska	65.1	67.2	70.0	2.9 *	1.12
Arizona	60.5	63.1	60.2	-2.9 *	1.12
Arkansas	33.7	36.3	48.4	12.1 *	1.34
Colorado	74.1	78.7	77.6	-1.0	0.94
Florida	58.2	60.8	64.1	3.4 *	0.92
Georgia	68.7	71.9	82.5	10.6 *	1.35
Indiana	70.2	72.2	71.5	-0.7	1.23
Iowa	75.6	77.6	75.6	-2.0	1.09
Louisiana	56.3	62.3	58.7	-3.6 *	1.37
Maryland	53.0	60.4	58.4	-2.0	1.22
Mississippi	52.5	54.4	53.6	-0.8	1.13
Nevada	51.1	49.6	53.8	4.1 *	0.89
New Jersey	63.9	67.8	68.5	0.7	1.10
New Mexico	23.6	28.5	29.7	1.2	0.92
North Dakota	65.5	70.6	68.0	-2.6 *	1.28
Ohio	62.7	65.0	74.0	9.0 *	1.22
Pennsylvania	62.4	68.9	69.7	0.8	1.16
South Carolina	23.8	25.7	19.9	-5.7 *	1.18
Tennessee	87.8	90.8	88.5	-2.3 *	0.91
Texas	60.9	67.4	71.9	4.5 *	0.99
Wisconsin	74.9	75.4	74.4	-1.0	1.16
2005 and 2007 state assessment reported results are not comparable					
Connecticut	75.9	77.3	80.6	3.3 *	0.99
Delaware	56.3	58.8	62.9	4.1 *	1.29
Hawaii	20.4	23.8	25.7	1.9 *	0.79
Idaho	69.8	71.4	72.2	0.8	0.97
Illinois	54.2	56.1	80.4	24.3 *	1.10
Kentucky	37.1	42.6	49.9	7.3 *	1.21
Maine	29.0	33.6	50.6	17.0 *	0.98
Massachusetts	41.6	49.1	47.1	-2.0	1.18
Michigan	61.4	60.7	68.6	8.0 *	1.29
Missouri	15.3	18.8	43.0	24.3 *	1.27
Montana	71.3	71.6	59.7	-11.9 *	1.18
New York	56.2	56.9	59.4	2.5 *	1.27
North Carolina	83.9	85.3	65.9	-19.3 *	0.98
Oklahoma	67.5	71.0	79.8	8.8 *	1.29
Oregon	65.3	66.5	73.2	6.7 *	1.26
Virginia	82.8	84.3	79.7	-4.6 *	1.08
West Virginia	70.6	71.9	71.6	-0.3	1.04
Wyoming	36.8	44.5	62.1	17.6 *	1.59

* Difference is statistically significant at $p < .05$.

[1] This matches the NAEP percentage meeting 2005 standard in 2005, by definition.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table C-5. Number of states according to the comparability of state-reported results between 2005 and 2007, by the statistical significance of the discrepancy between NAEP and state measures of gains in grades 4 and 8 reading

	Grade 4 2005 and 2007 state assessment reported results		Grade 8 2005 and 2007 state assessment reported results	
Difference D	**Comparable**	**Not comparable**	**Comparable**	**Not comparable**
Not statistically significant	11	2	9	2
Statistically significant	11	10	15	12
Total	22	12	24	14

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table C-6. Number of states according to the comparability of state-reported results between 2005 and 2007, by the statistical significance of the discrepancy between NAEP and state measures of gains in grades 4 and 8 mathematics

	Grade 4 2005 and 2007 state assessment reported results		Grade 8 2005 and 2007 state assessment reported results	
Difference D	**Comparable**	**Not comparable**	**Comparable**	**Not comparable**
Not statistically significant	13	2	9	3
Statistically significant	8	12	12	15
Total	21	14	21	18

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table C-7. Selected changes to state reading assessments between 2005 and 2007, by whether reports of grade 4 reading achievement changes from 2005 to 2007 in the state test and NAEP agree, by state

State/jurisdiction	Changed cut scores	Changed time of administration	Changed assessment items	Used different assessment	Changed content standards	Changed proficiency standards	Changed accommodation policy	Changed test contractors	No significant changes
Changes in student achievement on the state test are not statistically significantly different from changes on NAEP									
Alaska									
California			√						
Colorado									√
Indiana									√
Iowa									
Maryland									√
Massachusetts		√	√						
Mississippi									√
New Mexico									
North Carolina	√				√	√			
Ohio									√
Changes in student achievement on the state test are statistically significantly larger than changes on NAEP									
Arkansas			√				√		
Louisiana				√					
North Dakota									√
South Carolina	√		√		√	√		√	
Texas	√		√						
Changes in student achievement on the state test are statistically significantly smaller than changes on NAEP									
Alabama			√						√
Florida								√	
New Jersey	√	√	√	√	√	√	√		
Tennessee									√
Washington									
Wisconsin									

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table C-8. Selected changes to state reading assessments between 2005 and 2007, by whether reports of grade 8 reading achievement changes from 2005 to 2007 in the state test and NAEP agree, by state

State/jurisdiction	Changed cut scores	Changed time of administration	Changed assessment items	Used different assessment	Changed content standards	Changed proficiency standards	Changed accommodation policy	Changed test contractors	No significant changes
Changes in student achievement on the state test are not statistically significantly different from changes on NAEP									
Arizona			√						
California			√						
Colorado									√
Indiana									√
Iowa									
New Jersey									
North Carolina									√
Ohio									√
Texas									√
Changes in student achievement on the state test are statistically significantly larger than changes on NAEP									
Alabama			√						√
Arkansas			√				√		
Florida								√	
Illinois			√	√				√	
Louisiana				√					
Nevada									
New Mexico									√
North Dakota									√
Pennsylvania	√		√		√	√		√	
Tennessee									√
Changes in student achievement on the state test are statistically significantly smaller than changes on NAEP									
Alaska									
Maryland									√
Mississippi									√
South Carolina									√
Wisconsin									

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Reading Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table C-9. Selected changes to state mathematics assessments between 2005 and 2007, by whether reports of grade 4 mathematics achievement changes from 2005 to 2007 in the state test and NAEP agree, by state

State/jurisdiction	Changed cut scores	Changed time of administration	Changed assessment items	Used different assessment	Changed content standards	Changed proficiency standards	Changed accommodation policy	Changed test contractors	No significant changes
Changes in student achievement on the state test are not statistically significantly different from changes on NAEP									
Alabama			√						√
Colorado									√
Florida								√	
Georgia	√		√		√	√		√	
Iowa									
Louisiana				√					
Massachusetts			√						
Mississippi									√
New Jersey									
New Mexico									√
South Carolina									√
Tennessee									√
Wisconsin									
Changes in student achievement on the state test are statistically significantly larger than changes on NAEP									
Alaska									
Arkansas			√				√		
California			√						
Maryland									√
Texas									√
Changes in student achievement on the state test are statistically significantly smaller than changes on NAEP									
Indiana									√
North Dakota									√
Washington									

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

Table C-10. Selected changes to state mathematics assessments between 2005 and 2007, by whether reports of grade 8 mathematics achievement changes from 2005 to 2007 in the state test and NAEP agree, by state

State/jurisdiction	Changed cut scores	Changed time of administration	Changed assessment items	Used different assessment	Changed content standards	Changed proficiency standards	Changed accommodation policy	Changed test contractors	No significant changes
Changes in student achievement on the state test are not statistically significantly different from changes on NAEP									
Colorado									√
Indiana									√
Iowa									
Maryland									√
Mississippi									√
New Jersey									
New Mexico									√
Pennsylvania	√		√		√	√		√	
Wisconsin									
Changes in student achievement on the state test are statistically significantly larger than changes on NAEP									
Alaska									
Arkansas			√				√		
Florida								√	
Georgia	√		√		√	√		√	
Nevada									
Ohio				√					
Texas									√
Changes in student achievement on the state test are statistically significantly smaller than changes on NAEP									
Arizona			√						
Louisiana				√					
North Dakota									√
South Carolina									√
Tennessee									√

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 and 2007 Mathematics Assessments. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, ED*Facts* SY 2006-07, Washington, DC, 2008. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2008.

www.ed.gov

ies.ed.gov