Comparison Between NAEP and State Mathematics Assessment Results: 2003

Volume 1

Research and Development Report

Comparison Between NAEP and State Mathematics Assessment Results: 2003

Volume 1

Research and Development Report

January 2008

Don McLaughlin
Victor Bandeira de Mello
Charles Blankenship
Kassandra Chaney
Phil Esra
Hiro Hikawa
Daniela Rojas
Paul William
Michael Wolman
American Institutes for Research

Taslima Rahman
Project Officer
National Center for Education Statistics

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

Suggested Citation

For ordering information on this report, write to

Content Contact

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

I n late January through early March of 2003, the National Assessment of Educational Progress (NAEP) grade 4 and 8 reading and mathematics assessments were administered to representative samples of students in approximately 100 public schools in each state. The results of these assessments were announced in November 2003. Each state also carried out its own reading and mathematics assessments in the 2002-2003 school year, most including grades 4 and 8. This report addresses the question of whether the results published by NAEP are comparable to the results published by individual state testing programs.

OBJECTIVES

Comparisons to address the following four questions are based purely on results of testing and do not compare the content of NAEP and state assessments.

- How do states' achievement standards compare with each other and with NAEP?

- Are NAEP and state assessment results correlated across schools?

- Do NAEP and state assessments agree on achievement trends over time?

- Do NAEP and state assessments agree on achievement gaps between subgroups?

How do states' achievement standards compare with each other and with NAEP?

Both NAEP and State Education Agencies have set achievement, or performance, standards for mathematics and have identified test score criteria for determining the percentages of students who meet the standards. Most states have multiple performance standards, and these can be categorized into a *primary standard*, which, since the passage of *No Child Left Behind*, is generally the standard used for reporting adequate yearly progress (AYP), and standards that are above or below the primary standard. Most states refer to their primary standard as *proficient* or *meets the standard*.

By matching percentages of students reported to be meeting state standards in schools participating in NAEP with the distribution of performance of students in those schools on NAEP, cutpoints on the NAEP scale can be identified that are equivalent to the scores required to meet a state's standards.

From the analyses presented in chapter 2, we find:

- The median of the states' primary mathematics standards, as reflected in their NAEP equivalents, is between the NAEP *basic* and *proficient* levels in both grades 4 and 8.

- The primary standards vary greatly in difficulty across states, as reflected in their NAEP equivalents. In fact, among states, there is more variation in placement of primary mathematics standards than in average NAEP performance.

- As a corollary, states with high primary standards tend to see few students meet their standards, while states with low primary standards tend to see most students meet their standards.

- There is no evidence that setting a higher state standard is correlated with higher performance on NAEP. Students in states with high primary standards score just about the same on NAEP as students in states with low primary standards.

Are NAEP and state assessment results correlated across schools?

An essential criterion for the comparison of NAEP and state assessment results in a state is that the two assessments agree on which schools are high achieving and which are not. The critical statistic for testing this criterion is the correlation between schools' percentages achieving their primary standard, as measured by NAEP and the state assessment. Generally, a correlation of at least .7 is important for confidence in linkages between them.[1] Several factors other than similarity of the assessments depress this correlation. In 2003, correlations between NAEP and state assessment measures of mathematics achievement were greater than .7 in 41 of 46 states for grade 8 and in 30 of 49 states for grade 4.

One of these factors is a disparity between the standards: the correlation between the percent of students meeting a high standard on one test and a low standard on the other test are bound to be lower than the correlation between percents of students meeting standards of equal difficulty on the two tests. To be fair and unbiased, comparisons of percentages meeting standards on two tests must be based on equivalent standards for both tests. To remove the bias of different standards, NAEP was rescored in terms of percentages meeting the state's standard. Nevertheless, as discussed in chapter 3, other factors also depressed the correlations:

- Correlations are biased downward by schools with small enrollments, by use of scores for an adjacent grade rather than the same grade, and by standards set near the extremes of a state's achievement distribution.

---

1. A correlation of at least .7 implies that 50% or more of the variance of one variable can be predicted from the other variable.

- Estimates of what the correlations would have been if they were all based on scores on non-extreme standards in the same grade in schools with 30 or more students per grade were greater than .7 in 42 of 43 states for grade 8 and in 37 of 46 states for grade 4.[2]

Do NAEP and state assessments agree on achievement trends over time?

Comparisons are made between NAEP and state assessment mathematics achievement trends between 2000 and 2003. Achievement trends are measured by both NAEP and state assessments as gains in school-level percentages meeting the state's primary standard.[3]

From the analyses presented in chapter 4, we find:

- For mathematics achievement trends from 2000 to 2003, there are significant differences between NAEP and state assessment trends in 14 of 24 states in grade 4 and 11 of 22 states in grade 8.

- In aggregate, in grade 4 but not in grade 8, mathematics achievement gains from 2000 to 2003 measured by NAEP are significantly larger than those measured by state assessments.

- Across states, there was a positive correlation between gains measured by NAEP and gains measured by state assessments (r = .52 at grade 4 and r = .36 at grade 8).

Do NAEP and state assessments agree on achievement gaps between subgroups?

Comparisons are made between NAEP and state assessment measurement of mathematics achievement gaps in grades 4 and 8 in 2003. Comparisons are based on school-level percentages of Black, Hispanic, White, and economically disadvantaged and non-disadvantaged students achieving the state's primary mathematics achievement standard in the NAEP schools in each state.

From the analyses presented in chapter 5, we find:

- In 34 of 70 gap comparisons at grade 4 and 17 of 62 gap comparisons at grade 8, NAEP found significantly larger gaps than the state assessment did. In only two of the comparisons (both at grade 8) did the state assessment record a significantly larger gap.

- The tendency for NAEP to find larger gaps in mathematics achievement than state assessments did was equally strong with respect to Black-White and Hispanic-White gaps and slightly weaker for poverty gap comparisons.

---

2. Three states for which state reports of percentages meeting standards were unavailable were not included in the computations of these estimates.
3. To provide an unbiased trend comparison, NAEP was rescored in terms of the percentages meeting the state's primary standard in the earliest trend year.

DATA SOURCES

This report makes use of test score data for 48 states and the District of Columbia from two sources: (1) NAEP plausible value files for the states participating in the 2000 and 2003 mathematics assessments, augmented by imputations of plausible values for the achievement of excluded students;[4] and (2) state assessment files of school-level statistics compiled in the National Longitudinal School-Level State Assessment Score Database (NLSLSASD).[5]

All comparisons in the report are based on NAEP and state assessment results in schools that participated in NAEP, weighted to represent the states. Across states in 2003, the median percentage of NAEP schools for which state assessment records were matched was greater than 99 percent. However, results in this report represent about 96 percent of the regular public school population, because for confidentiality reasons state assessment scores are not available for the smallest schools in most states.

In most states, comparisons with NAEP grade 4 and 8 results are based on state assessment scores for the same grades, but in a few states for which tests were not given in grades 4 and 8, assessment scores from adjacent grades are used.

Because NAEP and state assessment scores were not available from all states prior to 2003, trends could not be compared in all states. Furthermore, in eight of the states with available scores, either assessments or performance standards were changed between 2000 and 2003, precluding trend analysis in those states for some years. As a result, comparisons of trends from 2000 to 2003 are possible in 24 states for grade 4 and 21 states for grade 8.

Because subpopulation achievement scores were not systematically acquired for the NLSLSASD prior to 2002, achievement gap comparisons are limited to gaps in 2003. In addition, subpopulation data are especially subject to suppression due to small sample sizes, so achievement gap comparisons are not possible for groups consisting of fewer than ten percent of the student population in a state.

Black-White gap comparisons for 2003 are possible in 25 states for grade 4 and 20 states for grade 8; Hispanic-White gap comparisons in 14 states for both grades 4 and 8; and poverty gap comparisons in 31 states for grade 4 and 28 states for grade 8.

---

4. 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/pubs/guide97/ques11.asp.

5. Most states have made school-level achievement statistics available on state web sites since the late 1990s; these data have been compiled into a single database, the NLSLSASD, for use by educational researchers. These data can be downloaded from http://www.schooldata.org. However, 2003 school-level state mathematics assessment results were not available for Nebraska and West Virginia when this report was prepared.

CAVEATS

Although this report brings together a large amount of information about NAEP and state assessments, there are significant limitations on the conclusions that can be reached from the results presented.

First, this report does not address questions about the content, format, or conduct of state assessments, as compared to NAEP. The only information presented in this report concerns the results of the testing—the achievement scores reported by NAEP and state mathematics assessments.

Second, this report does not represent all public school students in each state. It does not represent students in home schooling, private schools, or many special education settings. State assessment scores based on alternative tests are not included in the report, and no adjustments for non-standard test administrations (i.e., accommodations) are applied to scores. Student exclusion and nonparticipation are statistically controlled for NAEP data, but not state assessment data.

Third, this report is based on school-level percentages of students, overall and in demographic subgroups, who meet standards. As such, it has nothing to say about measurement of individual student variation in achievement within these groups or differences in achievement that fall within the same discrete achievement level.

Finally, this report is not an evaluation of state assessments. State assessments and NAEP are designed for different, although overlapping purposes. In particular, state assessments are designed to provide important information about individual students to their parents and teachers, while NAEP is designed for summary assessment at the state and national level. Findings of different standards, different trends, and different gaps are presented without suggestion that they be considered as deficiencies either in state assessments or in NAEP.

CONCLUSION

There are many technical reasons for different assessment results from different assessments of the same skill domain. The analyses in this report have been designed to eliminate some of these reasons, by (1) comparing NAEP and state results in terms of the same performance standards, (2) basing the comparisons on scores in the same schools, and (3) removing the effects of NAEP exclusions on trends. However, other differences remain untested, due to limitations on available data.

The findings in this report must necessarily raise more questions than they answer. For each state in which the correlation between NAEP and state assessment results is not high, a variety of alternative explanations must be investigated before reaching conclusions about the cause of the relatively low correlation. The report evaluates some explanations but leaves others to be explained when more data become available.

Similarly, the explanations of differences in trends in some states may involve differences in populations tested, differences in testing accommodations, or other technical differences, even though the assessments may be testing the same domain of skills. Only further study will yield explanations of differences in measurement of achievement gaps. This report lays a foundation for beginning to study the effects of differences between NAEP and state assessments of mathematics achievement.

Acknowledgments

There are many people to thank for the development of this report. First, thanks go to the people who made the state data available, including state testing programs, school administrators, and students. Second, thanks go to NCES and, in particular to the NAEP team, for working closely with the authors and painstakingly reviewing multiple drafts of this report. They include Peggy Carr, Steve Gorman, Taslima Rahman, and Alex Sedlacek. From NCES the reviewers included Jack Buckley, Patrick Gonzales, and Mike Ross. Lisa Bridges, from the Institute of Education Sciences, conducted the final review of the report and provided valuable feedback. Also, thanks go to the ESSI review team with Young Chun, John Clement, Kim Gattis, Rita Joyner, Lawrence Lanahan, and Michael Planty. Special thanks also go to Karen Manship and Jill Crowley, from AIR, who contributed to the development of this publication and coordinated the review process.

The authors are grateful for the assistance of the following persons who reviewed the report at different stages of development and made valuable comments and suggestions. Reviewers from outside the U.S. Department of Education were Pat DeVito, Assessment and Evaluation Concepts, Inc.; Debra Kalenski, Touchstone Applied Science Associates, Inc.; Lou Frabrizio, North Carolina Department of Education; Ed Haertel, Stanford University; Bob Linn, University of Colorado at Boulder; Wendy Yen, Education Testing Services; and Rick White, Kate Laguarda, Leslie Anderson, and Gina Hewes, Policy Studies Associates.

Contents

List of Tables

Table   **Page**

Table   **Page**

List of Figures

Figure   **Page**

1 Introduction

Achievement testing has a long history in American schools, although until the past 30 years its primary focus was on individual students, for either diagnostic or selection purposes. This began to change in the 1960s, with the increased focus on ensuring equality of educational opportunities for children of racial/ethnic minorities and children with special needs. In the 1970s, the U.S. government funded the National Assessment of Educational Progress (NAEP), whose mission was to determine, over the course of ensuing years and decades, how America was doing at reducing achievement gaps and improving the achievement of all students.[1]

For more than 30 years, NAEP has continued as an ongoing congressionally-mandated survey designed to measure what students know and can do. The goal of NAEP is to estimate educational achievement and changes in that achievement over time for American students of specified grades as well as for subpopulations defined by demographic characteristics and by specific background characteristics and experiences.

Calls for school reform in the 1980s and 1990s focused national attention on finding ways for schools to become more effective at improving the reading and mathematics achievement of their students. In 1990, state governors agreed on challenging goals for academic achievement in public schools by the year 2000.[2] School accountability for student reading and mathematics achievement reached a significant milestone in 2001 with the passage of the *No Child Left Behind Act*, which sets forth the goal that all students should be proficient in reading and mathematics by 2013-14.

No Child Left Behind created regulations and guidelines for measuring *Adequate Yearly Progress* (AYP). State education agencies report each year on which schools meet their AYP goals and which are in need of improvement. The determination of whether a school meets its goals involves a complex series of decisions in each state as to what criteria to use, what exclusions to authorize, and how to interpret the results. NAEP, on the other hand, does not report on AYP for schools; therefore, this report

---

1. On the history of the federal involvement in education and the creation of a national student assessment system during the 1960s, see Vinovskis (1998). For a historical perspective on testing and accountability see Ravitch (2004). For general information about NAEP, see the *NAEP Overview* at http://nces.ed.gov/nationsreportcard/about/.
2. *Goals 2000: Educate America Act*: http://www.ed.gov/legislation/GOALS2000/TheAct/sec102.html

will not address questions about states' compliance with *No Child Left Behind* requirements.

In January through March of 2003, NAEP grade 4 and 8 reading and mathematics assessments were administered to representative samples of students in approximately 100 public schools in each state. The results of these assessments were announced in November 2003. Each state also carried out its own reading and mathematics assessments in the 2002-2003 school year, most including grades 4 and 8. Many people are interested in knowing whether the results published by NAEP are the same as the results published by individual state testing programs. In this report, our aim is to construct and display the comparisons for mathematics achievement in a valid, reliable, fair manner. A companion report focuses on comparisons for reading achievement.

Although this report does not focus on AYP measurement specifically, it does focus on the measurement of mathematics achievement, and specifically on comparisons of the messages conveyed by NAEP mathematics assessment results and state mathematics assessment results.

These comparisons center on four facets of NAEP and state assessment results:

- Achievement standards
- School-level achievement percentages
- Achievement trends
- Achievement gaps

These facets of comparisons are summarized below.

COMPARING STATE MATHEMATICS ACHIEVEMENT STANDARDS

In recent years, states have expressed the achievement of the students in their schools in terms of the percentage who are meeting specified performance standards, similar in concept to NAEP's basic, proficient, and advanced achievement levels. Because each state's standards are set independently, the standards in different states can be quite different, even though they may have the same name. Thus, a student whose score is in the *proficient* range in one state can move to another state and find that his knowledge and skills produce a score that is *not* in the *proficient* range in that state. It would appear at first to be impossible to tell whether being proficient (i.e., meeting the proficiency standard) in one state is harder than being proficient in another state without having either some students take both states' tests or students in both states take the same test. However, NAEP can provide the needed link if results of the two states' tests are each sufficiently correlated with NAEP results.

COMPARING SCHOOLS' PERCENTAGES MEETING STATE PERFORMANCE STANDARDS

State assessment programs report the percentages of each school's students who achieve the state mathematics standards, and an important question is the extent to which NAEP and the state assessments agree on the ranks of the schools.[3] The critical statistic for measuring that agreement is the correlation between NAEP and state assessment results for schools. The question of how strongly NAEP and state mathematics assessment results are correlated is basic to the comparison of these two types of assessment. If they are strongly correlated, then one can expect that if NAEP had been administered in all schools in a state, the results would mirror the observed variations among schools' state assessment scores. Unfortunately, a variety of factors can lead to low correlations between tests covering the same content.

First, since the comparison is between percentages of students meeting standards, differences in the positions of those standards in the range of student achievement in a state will limit the correlations. Correlation between the percent of students meeting a high standard on one test and a low standard on another test will likely be substantially less than the correlation between two standards at the same position. This distortion in measuring the correlation between NAEP and state assessment results in a state is removed by *scoring* NAEP in terms of the percent meeting the equivalent of that state's standard.

This report explores three non-content factors that tend to depress correlations:

- differences in grade tested (the state may test in grades 3, 5, 7, or 9, instead of grade 4 or 8);

- small numbers of students tested (by NAEP or the state assessment) in some small schools, yielding less stable percentages of students meeting standards in each school;

- extremely high (or extremely low) standards.

This third factor yields very low (or very high) percentages meeting the standard across nearly all schools in the state, restricting the reliable measurement of differences among schools. Other potential non-content factors that may depress correlations include differences in accommodations provided to students with disabilities and English language learners, differences in motivational contexts, and time of year of testing.

---

3. Although NAEP's sample design does not generate school-level statistics that are sufficiently reliable to justify publication, state summaries of distributions of school-level statistics are appropriate for analysis.

COMPARING ACHIEVEMENT TRENDS AS MEASURED BY NAEP AND STATE ASSESSMENTS

Of central concern to both state and NAEP assessment programs is the comparison of achievement trends over time (e.g., USDE 2002, NAGB 2001). The extent to which NAEP measures of achievement trends match states' measures of achievement trends may be of interest to state assessment programs, the federal government, and the public in general.

Unlike state assessments, NAEP is not administered every year, and NAEP is only administered to a sample of students in a representative sample of schools in each state. For this report, the comparison of trends in mathematics achievement is limited to changes in achievement between the 1999-2000 school year and the 2002-2003 school year (i.e., between 2000 and 2003). For research purposes, analysts may wish to examine trends in earlier NAEP mathematics assessments (in 1992 and 1996), but matched state assessment data are not sufficiently available for those early years to warrant inclusion in this report.

COMPARING ACHIEVEMENT GAPS AS MEASURED BY NAEP AND STATE ASSESSMENTS

A primary objective of federal involvement in education is to ensure equal educational opportunity for all children, including minority groups and those living in poverty (USDE 2002). NAEP has shown that although there have been gains since 1970, certain minority groups lag behind other students in mathematics achievement in both elementary and secondary grades (Campbell, Hombo, and Mazzeo, 2000). There are numerous programs across the nation aimed at reducing the achievement gap between minority students and other students, as well as between schools with high percentages of economically disadvantaged students and other schools;[4] and state assessments are monitoring achievement to determine whether, in their state, the gaps are closing. This report addresses the specific research question:

> *Does NAEP's measurement of the grades 4 and 8 mathematics achievement gaps in each state in 2002-2003 differ from the state assessment's measurement of the same gaps?*

In future reports, it will be possible to compare trends in mathematics gaps between successive NAEP assessments, but data for such a comparison are not available for this report.

---

4. The poverty gap in achievement refers to the difference in achievement between economically disadvantaged students and other students, where disadvantaged students are defined as those eligible for free or reduced-price lunch.

SUPPORTING STATISTICS

Among the sources of differences in trends and gaps are sampling variation and variations in policies for accommodating and excluding students with disabilities and English language learners. Statistics bearing on these factors are included in this report as an aid for interpreting trends and gaps. Finally, this report assesses the impact of NAEP sampling by comparing state assessment results based on the NAEP schools with state assessment results reported on the state web sites.[5]

Some of the students with disabilities and English language learners selected to participate in NAEP are excused, or excluded, because it is judged that it would be inappropriate to place them in the test setting. NAEP's reports of state trends and comparisons of subgroup performance in the Nation's Report Card are based on standard NAEP data files, which are designed to represent the (sub)population of students in a state who would not be excluded from participation if selected by NAEP. In some cases, these trends are different from the trends that would have been reported if the excluded students had been included. To provide a firm basis for comparing NAEP and state assessment results, NAEP results presented in this report are based on full population estimates. These estimates extend the standard NAEP data files used in producing the Nation's Report Card by including representation of the achievement of the subset of the students with disabilities and English language learners who are excluded by NAEP. Corresponding results based on the standard NAEP estimates are presented in appendix C.

CAVEATS

This report does *not* address questions about the content, format, or conduct of state assessments, as compared to NAEP. The only information presented in this report concerns the results of the testing—the achievement scores reported by NAEP and state mathematics assessments. Although finding that the correlation between NAEP and state assessment results is high suggests that they are measuring similar skills, the only inference that can be made with assurance is that the schools where students achieve high NAEP mathematics scores are the same schools where students achieve high state assessment mathematics scores. It is conceivable that NAEP and the state assessment focus on different aspects of the same skill domain, but that the results are correlated because students master the different aspects of the domain together.

This report does *not* necessarily represent all students in each state. It is based only on NAEP and state assessment scores in schools that participated in NAEP. Although the results use NAEP weights to represent regular public schools in each state, they do not represent students in home schooling, private schools, or special education

---

5. Links to these web sites can be found at http://www.schooldata.org/, along with details regarding timing, publishers, and history of state tests.

settings not included in the NAEP school sampling frame. NAEP results are for grades 4 and 8, and they are compared to state assessment results for the same grade, an adjacent grade, or a combination of grades. State assessment scores based on alternative tests are not included in the report, and no adjustments for non-standard test administrations (*accommodations*) are applied to scores. Student exclusion and nonparticipation are statistically controlled for in NAEP data, but not for state assessment data.

This report does *not* address questions about NAEP and state assessment of individual variation of students' mathematics achievement within demographic groups within schools. The only comparisons in this report are between NAEP and state assessments of school-level scores, in total and for demographic subgroups. This is especially important in interpreting the measurement of achievement gaps, because the comparisons are blind to the variation of achievement within demographic groups within schools. Information about the average achievement of, for example, Black students in a school does not tell us anything about the variation between the highest and lowest achieving Black students in that school. The implication of this limitation is that, although the average achievement gaps between, for example, Black and White students are accurately estimated, the overlap of Black and White school-level averages is less than the overlap of Black and White individual student scores.

For most states, this report does *not* address comparisons of average test scores. The only comparisons in this report are between percentages of students meeting mathematics standards, as measured by NAEP and state assessments.[6] However, comparisons between percentages meeting different standards on two different tests (e.g., *proficient* as defined by NAEP and *proficient* as defined by the state assessment) are meaningless, because they only serve to compare the results of the two assessment programs' standard-setting methodologies. In order to provide meaningful comparisons, it is necessary to compare percentages meeting the same standard, measured separately by NAEP and state assessments. Specifically, we identified the NAEP scale equivalent of each state mathematics standard and *rescored* NAEP in terms of the percentage meeting the equivalent of that state's standard.[7] All comparisons of achievement trends and gaps in this report are based on the states' standards, not on the NAEP achievement levels.

Finally, this report is not an evaluation of state assessments. State assessments and NAEP are designed for different, although overlapping purposes. In particular, state assessments are designed to provide important information about individual students to their parents and teachers, while NAEP is designed for summary assessment at the state and national level. They may or may not be focusing on the same aspects of

---

6. There is an exception: in the three states for which state reports of percentages meeting standards were unavailable—Alabama, Tennessee, and Utah—comparisons were of school-level medians of percentile scores.

7. Appendix A includes details on estimating the placement of state achievement standards on the NAEP scale.

mathematics achievement. Findings of different standards, different trends, and different gaps are presented without suggestion that they be considered as deficiencies either in state assessments or in NAEP.

DATA SOURCES

This report makes use of data from two categories of sources: (1) NAEP data files for the states participating in the 2000 and 2003 mathematics assessments, and (2) state assessment files of school-level statistics compiled in the National Longitudinal School-Level State Assessment Score Database (NLSLSASD).

NAEP statistics

The basic NAEP files used for this report are based on administration of test instruments to approximately 2,000 to 2,500 students, in approximately 100 randomly selected public schools, in each state and grade. The files include achievement measures and indicators of race/ethnicity, gender, disability and English learner status, and free-lunch eligibility for each selected student. Because state assessment data are only available at the school level, as an initial step in the analysis, NAEP data are aggregated to the school level for Black, White, Hispanic, economically disadvantaged and non-disadvantaged, and all students by computing the weighted means for NAEP students in each school. These school-level statistics are used to compute state-level summaries that are displayed and compared to state assessment results in this report. The database includes weights for each school to provide the basis for estimating state-level summaries from the sample.

Aggregation of highly unstable individual results to produce reliable summary statistics is a standard statistical procedure. All NAEP estimates in the Nation's Report Card are derived from highly unstable individual student results for students selected to participate in NAEP. At the individual student level, there is no question that NAEP results are highly unstable. However, NCES uses these highly unstable results to produce and publish reliable state-level summary statistics. This act of aggregating a set of highly unstable estimates into a single summary statistic *creates* the stability needed to support the publication of the state level results.[8]

This report also tabulates reliable state level summary statistics, based on the aggregation of highly unstable individual NAEP plausible values. As an intermediate step, this report first aggregates the highly unstable individual plausible values into somewhat less highly unstable school level results, then aggregates the school-level

---

8. NAEP results are based on a sample of student populations of interest. By design, NAEP does not produce individual scores since individuals are administered too few items to allow precise estimates of their ability. In order to account for such situations, NAEP uses plausible values, i.e., random draws of an estimated distribution of a student's ability–an empirically derived distribution of proficiency values that are conditional on the observed values of the test items and the student's background characteristics. Plausible values are then used to estimate population characteristics. Additional information is available at http://am.air.org and at the *NAEP Technical Documentation Website* at http://nces.ed.gov/nationsreportcard/tdw/.

results to produce reliable state-level summaries. The reason for the two-stage aggregation is that it enables pairing NAEP results at the school level to state assessment results in the same schools. The level of resulting stability of state level summary statistics is similar to the stability of state level results published in other NAEP reports.

NAEP estimates a distribution of possible (or *plausible*) values on an achievement scale for each student profile of responses on the assessment, producing an analysis file with five randomly selected achievement scale values consistent with the profile of responses. The NAEP mathematics achievement scale has a mean of approximately 230 in grade 4 and 275 in grade 8, with standard deviations of approximately 30 and 35 points, respectively. In this context, the random variation of imputed plausible values for each student profile, approximately 10 points on this scale, is too large to allow reporting of individual results, but the plausible values are appropriate for generating state-level summary statistics. Standards for *basic*, *proficient*, and *advanced* performance are equated to *cutpoints* on the achievement scale. Details of the NAEP data are described at http://nces.ed.gov/nationsreportcard.

On NAEP data files used for the Nation's Report Card (referred to as standard NAEP estimates), achievement measures are missing for some students with disabilities and English language learners, as noted above. These excluded students represent roughly four percent of the student population.[9] In order to avoid confounding trend and gap comparisons with fluctuating exclusion rates, NAEP reported data have been extended for this report to include imputed plausible values for students selected for NAEP but excluded because they are students with disabilities or English language learners who were deemed unable to participate meaningfully in the NAEP assessment. We refer to the statistics including this final four percent of the selected population as *full population statistics*, as distinguished from the reported data used in the Nation's Report Card. The methodology used to estimate the performance of excluded students makes use of special questionnaire information collected about all students with disabilities and English language learners selected for NAEP, whether they completed the assessment or not, and is described in Appendix A and by McLaughlin (2000, 2001, 2003) and is validated by Wise, et al. (2004).

State assessment school-level statistics

Most states have made school-level achievement statistics available on state web sites since the late 1990s; these data have been compiled by the American Institutes for Research for the U.S. Department of Education into a single database, the NLSLSASD, for use by educational researchers. These data can be downloaded from http://www.schooldata.org.

The NLSLSASD contains scores for over 80,000 public schools in the country, in most cases for all years since 1997-98. These scores are reported separately by each

---

9. The average percentage excluded for all states in 2003 is close to four percent at both grades. The exclusion rates vary between states, and within states, between years.

state for each subject and grade. In most cases, multiple measures are included in the database for each state test, such as average scale scores and percentages of students meeting state standards; for a few states, multiple tests are reported in some years. Starting in the 2001-2002 school year, the NLSLSASD added subpopulation breakdowns of school-level test scores reported by states.

Three factors limit our use of these data for this report. First, the kind of score reported changes from time to time in some states. For uses of these scores that compare some schools in a state to other schools in the same state, the change of scoring is not a crucial limitation; however, for measurement of whole-state achievement trends, which is a central topic of this report, changes in tests, standards, or scoring create a barrier for analysis. Discrepancies between NAEP and state assessment reports of mathematics achievement trends may, for some states, merely reflect state assessment instrumentation or scoring changes.

Second, not all states reported mathematics achievement scores for grades 4 and 8 in 2002-2003. Because mathematics achievement is cumulative over the years of elementary and secondary schooling, the mathematics achievement scores for different grades in a school are normally highly correlated with each other. Therefore, NAEP grade 4 trends can be compared to state assessment grade 3 or grade 5 trends, and NAEP grade 8 trends can be compared to state assessment grade 7 or grade 9 trends.[10] More discrepancies between NAEP and state assessment results are to be expected when they are based on adjacent grades, not the same grade, primarily because the same grade comparisons are between scores of many of the same students while adjacent grade comparisons involve different cohorts of students. The magnitude of this effect is described in the section on correlations.

Third, the state achievement information on subpopulations is only available for 2003, so NAEP and state assessment reports of trends in gaps cannot be compared in this report. Also, because the NLSLSASD makes use of information available to the public, the scores for very small samples are suppressed. Thus, schools with state assessment scores on fewer than a specified number of students in a subpopulation (e.g., 5) are excluded from the analysis. The *suppression* threshold varies among the states. The suppression threshold is included in the description of each state's assessment in the State Profiles section of this report (Appendix D).

Each state has set standards for achievement, and recent public reports include percentages of students in each school meeting the standards. Most states report percentages for more than one level, and they frequently report the percentages at each level.[11] In this report, percentages meeting standards are always reported as the percentages *at* or *above* a level. For example, if a state reports in terms of four levels (based on three standards), and a school is reported to have 25 percent at each level, this report will indicate that 75 percent met the first standard, 50 percent met the

10. Comparison of NAEP grade 8 scores with state assessment grade 9 scores is only possible in some states, because in other states, very few schools serve both grades.

11. Five states reported only a single level: Iowa, Nebraska, North Dakota, Rhode Island, and Texas.

second standard, and 25 percent met the third (highest) standard. Some states also make available median percentile ranks, average scale scores, and other school-level statistics. For uniformity, when available, the analyses in this report will focus on percentages of students meeting state standards.[12] These percentages may not exactly match state reports because they are based on the NAEP representative sample of schools.

Sample sizes and percentages of the NAEP samples used in comparisons are shown in table 1. The number of public schools selected for NAEP in each state is shown in the first column, and the number of these schools included in the comparisons in this report is shown in the second column. The percent of the student population represented by the comparison schools is shown in the third column. (Table 20, later in the report, shows the percentage of schools that were able to be matched with usable assessment score data.)

The percentages of the population represented by NAEP used in the comparisons are less than 100 percent where state assessment scores are missing for some schools.[13] They may be missing either because of failure to match schools in the two surveys or because scores for the school are suppressed on the state web site (because they are based on too few students). Because the schools missing state assessment scores are generally small schools, percentages of student populations represented by the schools used in the comparisons are generally higher than the percentages of schools. The most extreme examples are Alaska and North Dakota: for Alaska the grade 8 comparisons are based on 57 percent of the NAEP schools, but these schools serve 90 percent of the students represented by NAEP; and for North Dakota the grade 8 comparisons are based on 21 percent of the NAEP schools, but these schools serve 62 percent of the students represented by NAEP.

Across states, the median percentage of the student population represented is 96 percent for grade 4 and 97 percent for grade 8. For individual states, the percentages included in comparisons are greater than 80 percent, with four exceptions: Delaware (58 percent for grade 4), New Mexico (74 percent for grade 4 and 71 percent for grade 8), and North Dakota (62 percent for grade 8). Grade 5 assessment scores were used for Delaware, and only 57 percent of the NAEP schools (representing 58 percent of the population) had grade 5 state assessment scores. The New Mexico and North Dakota exceptions are based on suppressed scores of small schools, since more than 90 percent of the NAEP schools were successfully matched to state assessment records in these states.

Not all states are included in this report. 2003 school-level mathematics state assessment results were not available for Nebraska. In Minnesota, New Hampshire,

---

12. All state assessment figures presented are percentages of students achieving a standard except for Alabama, Tennessee, and Utah, for which only median percentile ranks are available.

13. A very small number of NAEP schools (fewer than one percent in most states) are also omitted due to lack of success in matching them to state assessment records. Rates of success in matching are described in the report section on supporting statistics.

and Ohio, the results were only available for elementary schools. For Minnesota, an attempt was made to match grade 9 state mathematics assessment scores to schools participating in the grade 8 NAEP assessment, but this failed because very few schools in Minnesota served both 8th and 9th grades. For West Virginia, the only state assessment results available were for a composite of reading and mathematics, and these results were included in the companion report, *Comparison Between NAEP and State Reading Assessment Results: 2003* (McLaughlin et al., 2008).

Table 1.   Number of NAEP schools, number of NAEP schools available for comparing state assessment results with NAEP results in grades 4 and 8 mathematics, and the percentage of the student population in these comparison schools, by state: 2003

State/ jurisdiction	Grade 4			Grade 8		
	NAEP schools	Comparison schools	Percent of population	NAEP schools	Comparison schools	Percent of population
Alabama	112	106	92.4	104	100	95.4
Alaska	154	110	89.6	100	57	89.9
Arizona	121	99	82.4	118	105	93.6
Arkansas	119	117	98.7	109	101	93.2
California	253	216	95.1	188	180	99.3
Colorado	124	111	91.9	114	104	97.2
Connecticut	110	108	98.5	104	102	97.8
Delaware	88	50	57.8	37	32	93.4
District of Columbia	118	99	86.4	38	26	82.3
Florida	106	103	97.4	97	96	99.1
Georgia	156	147	95.6	117	113	95.4
Hawaii	107	107	100.0	67	54	97.0
Idaho	124	114	96.3	91	86	97.4
Illinois	174	161	89.5	170	169	99.2
Indiana	111	110	99.1	99	99	100.0
Iowa	136	133	98.6	116	114	98.5
Kansas	137	130	95.9	126	120	95.5
Kentucky	121	117	96.6	113	112	98.9
Louisiana	110	109	98.5	96	94	98.5
Maine	151	145	98.5	108	105	97.6
Maryland	108	106	97.3	96	96	100.0
Massachusetts	165	161	98.8	132	128	97.8
Michigan	136	133	98.5	111	105	96.7
Minnesota	113	100	88.2	—	—	—
Mississippi	111	107	95.8	108	102	89.8
Missouri	126	126	100.0	117	114	98.7
Montana	180	142	93.9	131	101	95.1
Nebraska	—	—	—	—	—	—
Nevada	111	107	96.7	67	63	95.7
New Hampshire	122	108	89.0	—	—	—
New Jersey	110	109	98.9	107	107	100.0
New Mexico	119	89	73.8	97	68	70.7
New York	149	145	97.2	148	141	95.4
North Carolina	153	151	99.6	132	129	97.5
North Dakota	209	176	94.0	144	31	62.1
Ohio	168	163	90.3	—	—	—
Oklahoma	137	132	95.7	129	123	96.8
Oregon	125	111	89.0	109	105	98.8
Pennsylvania	114	101	87.5	103	101	98.4
Rhode Island	114	111	99.2	54	51	97.9
South Carolina	106	101	96.7	98	92	93.2
South Dakota	187	143	90.7	137	106	92.9
Tennessee	116	96	81.2	108	94	83.1
Texas	197	194	97.0	146	142	95.8
Utah	113	104	91.5	94	91	96.9
Vermont	178	154	92.3	104	99	96.6
Virginia	116	107	90.6	107	103	94.0
Washington	109	96	88.6	103	85	83.3
West Virginia	—	—	—	—	—	—
Wisconsin	127	127	100.0	105	103	98.8
Wyoming	170	145	97.4	89	74	98.2

— Not available.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

2 Comparing State Performance Standards

Each state has set either one or several *standards* for performance in each grade on its mathematics assessment. We endeavored to select the primary standard for each state as the standard it uses for reporting *adequate yearly progress* to the public. However, we cannot be certain of success in all cases because in some states policies for reporting adequate yearly progress have changed. Short versions of the states' names for the standards are shown in table 2, with the primary standard listed as *standard 3*. NAEP has set three such standards, labeled *basic*, *proficient*, and *advanced*.

These standards are described in words, and they are operationalized as test scores above a corresponding *cutpoint*. This is possible for NAEP, even though the design of NAEP does not support reporting individual scores—NAEP is only intended to provide reliably reportable statistics for broad demographic groups (e.g., gender and racial/ethnic) at the state level or for very large districts.

Because each state's standards are set independently, the standards in different states can be quite different, even though they are named identically. Thus, a score in the *proficient* range in one state may not be in the *proficient* range in another state. Because NAEP is administered to a representative sample of public school students in each state, NAEP can provide the link needed to estimate the difference between two states' achievement standards.

The objective of this comparison is to place all states' mathematics performance standards for grades 4 and 8, or adjacent grades, on a common scale, along with the NAEP achievement levels. This comparison is valuable for two reasons. First, it sheds light on the variations between states in the percentages of students reported to be *proficient*, *meeting the standard*, or *making satisfactory progress*. Second, for comparisons between NAEP and state assessment trends and gaps, it makes possible the removal of one important source of bias: a difference between two years or between two subpopulations in percentages achieving a standard is affected as much by the choice of where that standard is set on the achievement scale as by instructional reform.

Table 2. Short names of state mathematics achievement performance standards, by state: 2003

State/ jurisdiction	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5
Alabama			Percentile Rank[1]		
Alaska		Below Proficient	Proficient	Advanced	
Arizona		Approaching	Meeting	Exceeding	
Arkansas		Basic	Proficient	Advanced	
California	Below Basic	Basic	Proficient	Advanced	
Colorado			Partially Proficient	Proficient	Advanced
Connecticut	Basic	Proficient	Goal	Advanced	
Delaware		Below	Meeting	Exceeding	Distinguished
District of Columbia		Basic	Proficient	Advanced	
Florida		Limited Success	Partial Success	Some Success	Success
Georgia			Meeting	Exceeding	
Hawaii		Approaching	Meeting	Exceeding	
Idaho		Basic	Proficient	Advanced	
Illinois	Starting	Approaching	Meeting	Exceeding	
Indiana			Pass	Pass Plus	
Iowa			Proficient		
Kansas		Basic	Proficient	Advanced	Exemplary
Kentucky		Apprentice	Proficient	Distinguished	
Louisiana	Approaching Basic	Basic	Mastery	Advanced	
Maine		Partially Meeting	Meeting	Exceeding	
Maryland			Proficient	Advanced	
Massachusetts	Warning	Needs Improvement	Proficient	Advanced	
Michigan		Basic	Meeting	Exceeding	
Minnesota	Partial Knowledge	Satisfactory	Proficient	Superior	
Mississippi		Basic	Proficient		
Missouri	Progressing	Nearing Proficient	Proficient	Advanced	
Montana		Nearing Proficient	Proficient	Advanced	
Nebraska			Meeting		
Nevada		Approaching	Meeting	Exceeding	
New Hampshire			Basic	Proficient	Advanced
New Jersey			Proficient	Advanced	
New Mexico		Top 75%	Top half	Top 25%	
New York		Need Help	Meeting	Exceeding	
North Carolina		Inconsistent Mastery	Consistent Mastery	Superior	
North Dakota			Meeting		
Ohio		Basic	Proficient	Advanced	
Oklahoma		Little Knowledge	Satisfactory	Advanced	
Oregon			Meeting	Exceeding	
Pennsylvania		Basic	Proficient	Advanced	
Rhode Island			Proficient		
South Carolina		Basic	Proficient	Advanced	
South Dakota		Basic	Proficient		
Tennessee			Percentile Rank		
Texas			Passing		
Utah			Percentile Rank		
Vermont	Below	Nearly	Achieved	Honors	
Virginia			Proficient	Advanced	
Washington		Below	Met	Above	
West Virginia		Top 75%	Top half	Top 25%	
Wisconsin		Basic	Proficient	Advanced	
Wyoming		Partially Proficient	Proficient	Advanced	

1. Percentile rank while not a standard, is needed for comparisons in Alabama, Tennessee, and Utah. Similarly, for New Mexico and West Virginia quartiles are used for comparisons.

NOTE: Standard 3 represents the primary standard for every state. In most cases, it is the criterion for Adequate Yearly Progress (AYP). The state standards listed above are those for which assessment data exist in the NLSLSASD.

SOURCE: The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

NAEP ACHIEVEMENT DISTRIBUTION

To understand the second point, we introduce the concept of a *population profile* of NAEP achievement. Achievement is a continuous process, and each individual student progresses at his or her own rate. When they are tested, these students demonstrate levels of achievement all along the continuum of mathematics skills, and these are translated by the testing into numerical scale values. Summarizing the achievement of a population as the percentage of students who meet a standard conveys some information, but it hides the profile of achievement in the population - how large the variation in achievement is, whether high-achieving students are few, with extreme achievement, or many, with more moderate achievement, and whether there are few or many students who lag behind the mainstream of achievement. A population profile is the display of the achievement of each percentile of the population, from the lowest to the highest, and by overlaying two population profiles, one can display comparisons of achievement gains and achievement gaps at each percentile. More important for the comparison of standards across states, a population profile can show how placement of a standard makes a difference in how an achievement gain translates into a gain in the percentage of students meeting that standard.

Figure 1 displays a population profile of mathematics achievement in grade 4, as measured by NAEP in 2003. To read the graph, imagine students lined up along the horizontal axis, sorted from the lowest performers on a mathematics achievement test at the left to the highest performers at the right. The graph gives the achievement score associated with each of these students. For reference, figure 1 also includes the NAEP scale scores that are thresholds for the achievement levels. The percentage of student scores at or above the basic threshold score of 214, for example (i.e., students who have achieved the basic level), is represented as the part of the distribution to the right of the point where the population profile crosses the basic threshold. For example, the curve crosses the basic achievement level at about the 24th percentile, which means that 24 percent of the student population scores below the basic level, while 76 percent scores at or above the basic level. Similarly, 32 percent of the population meets the proficient standard (scores at or above 249), and 5 percent of the population meets the advanced standard (scores at or above 282).

- The scale of achievement is the NAEP scale, ranging from 0 to 500; achievement ranges from less than 200 in the lowest 10 percent of the population to above 275, in the top 10 percent of the population.

- In the middle range of the population, from the 20th percentile to the 80th percentile, each percent of the population averages about 1 point on the NAEP scale higher than the next lower percent. At the extremes, where the slopes of the curve are steeper, the variation in achievement between adjacent percentages of the population is much greater.

Figure 1.   Distribution of NAEP mathematics scale scores for the nation's public school students at grade 4, with NAEP basic, proficient, and advanced thresholds: 2003
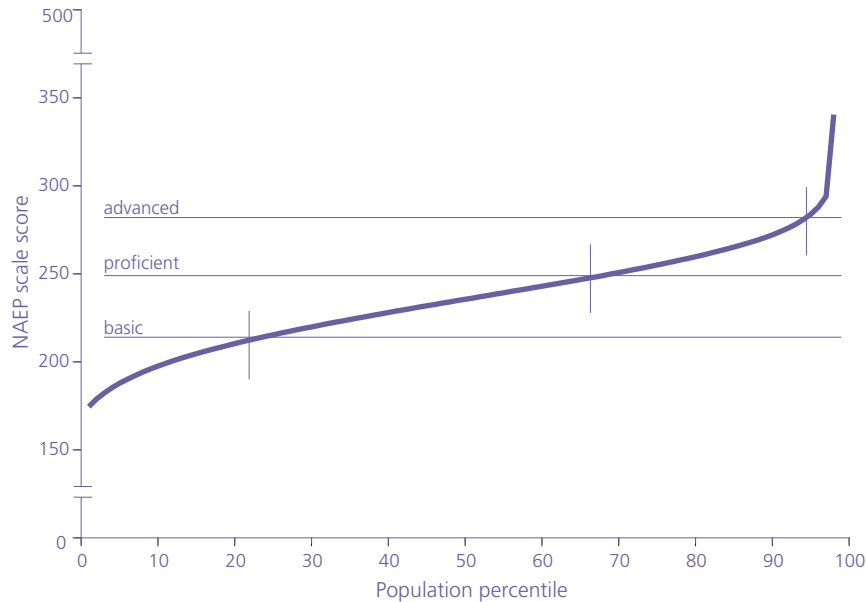
NOTE: Each point on the curve is the expected scale score for the specified percentile of the student population.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates.

Next we suppose that as a result of educational reform, everybody's mathematics achievement improves by 10 points on the NAEP scale. We can superimpose this hypothetical result on the population profile in figure 1, creating the comparison profile in figure 2. At each percentile of the population, the score in the hypothetical future is 10 points higher than in 2003. In the middle of the distribution, this is equivalent to a gain of about 13 percentile points (e.g., a student at the median in the future would be achieving at a level achieved by the 63rd percentile of students in 2003, or in other words, 50 percent of the future population would be achieving at levels only reached by 37 percent of students in 2003). Again, the NAEP basic, proficient, and advanced achievement thresholds are superimposed on the population profile.

As expected, the hypothetical profile of future achievement crosses the achievement thresholds at different points on the achievement continuum. In terms of percentages of students meeting standards, an additional 9 percent are above the basic cutpoint and an additional 13 percent are above the proficient cutpoint, but only 5 percent more are above the advanced cutpoint. Where the standard is set determines the gain in the percentage of the population reported to be achieving the standard. Percentage gains would appear to be twice as large for standards set in the middle of the distribution as for standards set in the tails of the distribution.

Figure 2.   Distribution of NAEP mathematics scale scores for the nation's public
school students at grade 4: 2003 and hypothetical future
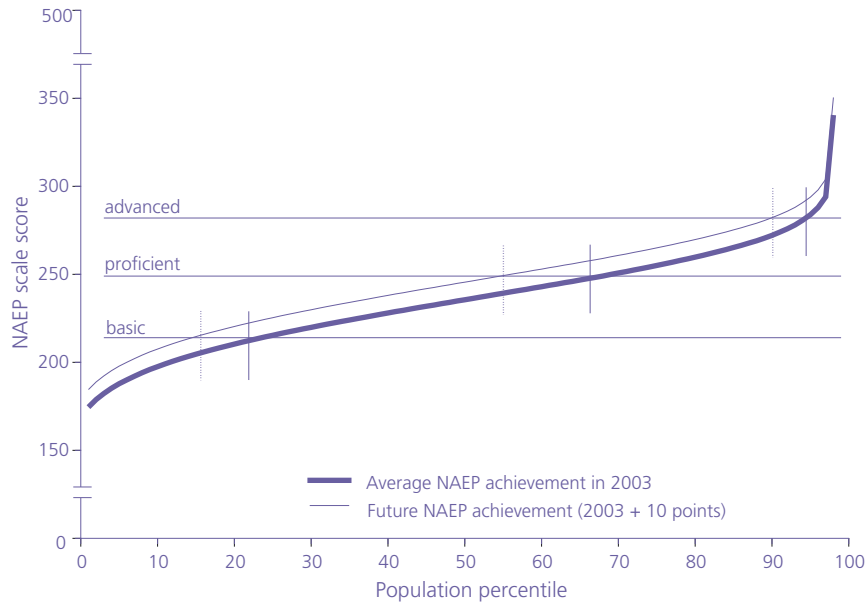
NOTE: Each point on the curve is the expected scale score for the specified percentile of the student population.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates.

This is important in comparing NAEP and state assessment results.[14] If NAEP's proficiency standard is set at a point in an individual state's distribution where achievement gains have small effects on the percentage meeting the standard, and if the state's proficiency standard is set at a point in the state's distribution where the same achievement gains have a relatively large effect on the percentage meeting the standard, then a simple comparison of percentages might find a discrepancy between NAEP and state assessment gains in percentages meeting standards when there is really no discrepancy in achievement gains.

The same problem affects measurement of gaps in achievement in terms of percentages meeting a standard. NAEP might find the poverty gap in a state to be larger than the state assessment reports merely due to differences in the positions of the state's and NAEP's proficiency standards relative to the state's population profiles for students in poverty and other students. And the problem is compounded in measurement of trends in gaps, or gap reduction.[15]

---

14.   Figure 1 is the distribution for the entire nation. The population profiles for individual states vary, although the NAEP cutpoints remain constant for all states.

15.   In this report, our interest is that variations in standards can distort comparisons between NAEP and state assessment gaps and trends. However, the same problem distorts comparisons of trends in percentages meeting standards between states.

The solution for implementing comparisons between NAEP and state assessment results is to make the comparisons at the same standard. This is possible if we can determine the point on the NAEP scale corresponding to the cutpoint for the state's standard. NAEP data can easily be re-scored in terms of any specified standard's cutpoint. The percentage of NAEP scale scores (*plausible values)* greater than the cutpoint is the percentage of the population meeting the standard.

NAEP Scale Equivalents

The method for determining the NAEP scale score corresponding to a state's standard is a straightforward equipercentile mapping. In nearly every public school participating in NAEP, a percentage of students meeting the state's achievement standard on its own assessment is also available. The percentage reported in the state assessment to be meeting the standard in each NAEP school is matched to the point in the NAEP achievement scale corresponding to that percentage. For example, if the state reports that 55 percent of the students in fourth grade in a school are meeting their achievement standard and 55 percent of the estimated NAEP achievement distribution in that school lies above 230 on the NAEP scale, then the best estimate from that school's results is that the state's standard is equivalent to 230 on the NAEP scale.[16] These results are aggregated over all of the NAEP schools in a state to provide an estimate of the NAEP scale equivalent of the state's threshold for its standard. The specific methodology is described in appendix A.

A strength and weakness of this method is that it can be applied to any set of numbers, whether or not they are meaningfully related. To ensure scores are comparable, after determining the NAEP scale equivalents for each state standard, we return to the results for each NAEP school and compute the discrepancy between (a) the percentage meeting the standard reported by the state for that school and (b) the percentage of students meeting the state standard estimated by NAEP data for that school. If the mapping were error-free, these would be in complete agreement; however, some discrepancies will arise from random variation. This discrepancy should not be noticeably larger than would be accounted for by simple random sampling variation. If it is noticeably larger than would be expected if NAEP and the state assessment were parallel tests, then we note that the validity of the mapping is questionable—that is, the mapping appears to apply differently in some schools than in others. As a criterion for questioning the validity of the placement of the state standard on the NAEP scale, we determine whether the discrepancies are sufficiently large to indicate that the NAEP and state achievement scales have less than 50 percent of variance in common.[17]

On the following pages, figures 3 and 4 display the NAEP scale score equivalents of primary grade 4 and grade 8 mathematics achievement standards in 45 states and the

---

16. The school's range of plausible achievement scale values for fourth grade is based on results for its NAEP sample of students.

District of Columbia.[18] In both grades the NAEP equivalents of the states' primary standards ranged from well below the NAEP basic level to slightly above the NAEP proficient level. In both grades, the median state primary standard was between the NAEP basic and proficiency thresholds.

The horizontal axis in figures 3 and 4 indicates the relative error criterion–the ratio of the errors in reproducing the percentages meeting standards in the schools based on the mapping to the size of errors expected by random measurement and sampling error if the two assessments were perfectly parallel. A value of 1.0 for this relative error is expected, and a value greater than 1.5 suggests that the mapping is questionable.[19] The numeric values of the NAEP scale score equivalents for the primary standards displayed in figures 3 and 4, as well as other standards, appear in tables B-1 and B-3 in Appendix B.

Eight of the 46 grade 4 mathematics standards have relative errors greater than 1.5, as indicated by their position to the right of the vertical line in the figure, and they are displayed in lower case letters in figure 3, indicating that the variation in results for individual schools was large enough to call into question the use of these equivalents. In six of these eight states (Delaware, Indiana, Kentucky, New Hampshire, Oregon, and Virginia), the state assessment results are for grade 3 or 5, so the comparison with NAEP fourth grade results is indirect. The grade discrepancy appears to be a more severe problem for mapping mathematics than for reading, possibly because the elementary school mathematics curriculum has more grade-specific learning objectives than reading. The problem with the mapping for Texas relates to a restriction of range: at 88 percent passing, it was the most extreme of the states' primary standards, leaving relatively little room for reliable measurement of achievement differences between schools. The other state for which the grade 4 mapping is questionable is Vermont.

17. This criterion is different from the usual standard error of equipercentile equating, which is related to the coarseness of the scales, not their correlation. With the relative error criterion we assessed the extent to which the error of the estimate is larger than it would be if NAEP and the state assessment were testing exactly the same underlying trait; in other words, by evaluating the accuracy with which each school's reported percentage of students meeting a state standard can be reproduced by applying the linkage to NAEP performance in that school. The method of estimation discussed in appendix A ensures that, on average, these percentages match, but there is no assurance that they match for each school. To the extent that NAEP and the state assessment are parallel, the percentages should agree for each school, but if NAEP and the state assessment are not correlated, then the mapping will not be able to reproduce the individual school results.

18. No percentages meeting mathematics achievement standards were available for this report for Alabama, Nebraska, Tennessee, Utah, and West Virginia.

19. The computation on which this distinction is made is described in appendix A.

Figure 3. NAEP scale equivalents of primary state mathematics achievement standards, grade 4 or adjacent grade, by relative error criterion: 2003
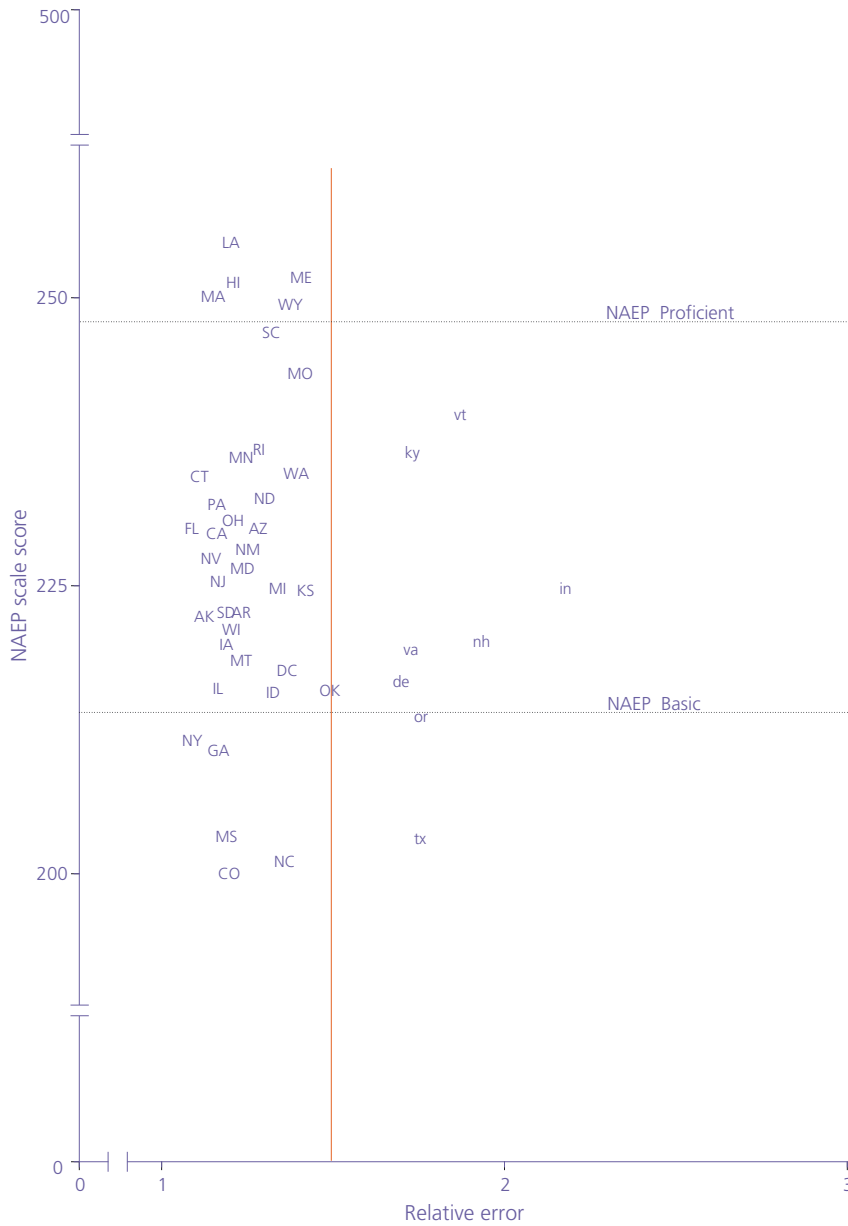
NOTE: Primary standard is the state's standard for *proficient* performance. Relative error is a ratio measure of reproducibility of school-level percentages meeting standards, described in appendix A. The vertical line indicates a criterion for maximum relative error. Standards for the eight states displayed in lowercase letters to the right of the vertical line have relative errors greater than 1.5; the variation in results for individual schools in these states is large enough to call into question the use of these equivalents.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 4.   NAEP scale equivalents of primary state mathematics achievement standards, grade 8 or adjacent grade, by relative error criterion: 2003

NOTE: Primary standard is the state's standard for *proficient* performance. Relative error is a ratio measure of reproducibility of school-level percentages meeting standards, described in appendix A. The vertical line indicates a criterion for maximum relative error.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Primary state standards for grade 8 mathematics are displayed in figure 4. For three states for which grade 4 data were available, we did not have data available for grade 8 comparisons.[20] The mappings for the remaining 43 states all appear to be acceptable.

Because this is an initial application of the relative error criterion for evaluating the validity of mapping state mathematics achievement standards to the NAEP scale, we have included the states for which our mappings are questionable in the comparison analyses. However, findings of differences between NAEP and state assessment results for trends and gaps should not be surprising given the quality of the mapping.

The thresholds for these primary state mathematics standards range from below the NAEP basic threshold (e.g., North Carolina and Georgia) to above the NAEP proficient threshold (e.g., Louisiana and Maine); this variation can have profound effects on the percentages of students states find to be meeting their standards. Focusing on the primary mathematics achievement standards, we can ask:

- How variable are the standards from one state to another?

- How is variability of standards related to the percentages of students meeting them?

- How is variation among standards related to the performance of students on NAEP?

In a broader arena, most states have set multiple standards, or achievement levels, and it may be of value to examine the variation in their placement of all levels in relation to the NAEP scale and to their student populations.

- Is there a pattern in the placement of standards relative to expected student performance?

These questions are addressed in the following pages.

How variable are the performance standards from one state to another?

In order to interpret information about the percentage of students meeting one state's standard and compare it to the percentages of students in other states meeting those other states' standards, it is essential to know how the standards relate to each other. Although many of the standards are clustered between the NAEP basic and proficient thresholds, there is great variability. The primary standards range from approximately the 15th to the 80th percentile of the NAEP mathematics achievement distribution. Thus it should not be surprising to find reports that in some states 70 percent of students are meeting the primary standard while 30 percent of students in other states are meeting their states' primary standards, but the students in the latter states score higher on NAEP. Such a result does not necessarily indicate

---

20. Grade 8 state mathematics assessment data were not available for Minnesota, New Hampshire, and Ohio.

that schools are teaching differently or that students are learning mathematics differently in the different states; it may only indicate variability in the outcomes of the standard setting procedures in the different states.

The variability of the NAEP scale equivalents of the states' primary standards is summarized in table 3. The standard deviations of 13.7 and 16.7 NAEP points among states' primary standards can be translated into the likelihood of finding contradictory assessment results in different states. To see this concretely, imagine a set of students who take one state's mathematics assessment and then another state's mathematics assessment. How different would the percentage of these students meeting the two states' standards be? In some pairs of states, with standards set at the same level of difficulty, we would expect only random variation, but in extreme cases, such as among fourth graders in Louisiana and North Carolina, the difference might be 50 percent (i.e., of a nationally representative sample of students, 50 percent more would appear to show *consistent mastery* in North Carolina than would appear to demonstrate *mastery* in Louisiana). On average, for any random pair of states, this discrepancy would be about 15 percentage points. That is, among sets of students in two randomly selected states who are actually achieving in mathematics at the same level, about 15 percent would be classified differently as to whether they were meeting the state's primary mathematics standard in the two states.

Table 3.　Mean and standard deviation of primary mathematics standard cutpoints across states, by grade: 2003

Level	Number of states	Average cutpoint	Standard error	Standard deviation	Standard error of standard deviation
Grade 4	46	227.0	0.16	13.7	0.17
Grade 8	43	229.3	0.18	16.7	0.18

NOTE: Primary standard is the state's standard for *proficient* performance.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

How is variability of performance standards related to the percentages of students meeting them?

Is it possible that states are setting standards in relation to their particular student populations, with higher standards set in states where mathematics achievement is higher? Perhaps one could imagine that public opinion might lead each state education agency to set a standard to bring all students up to the level currently achieved by the median student in its state. Then variation in standards would just be a mirror of variation in average achievement among the states. If that is not the case, then we should expect to see a negative relationship between the placement of the standard on the NAEP scale and the percentage of students meeting the standard.

This question is addressed in figures 5 and 6, which graph the relations between the difficulty of meeting each standard, as measured by its NAEP scale equivalent, and

the percentage of students meeting the standard. The higher the standard is placed, the smaller the percentage of students in the state meeting the standard. In fact, the negative relation is so strong that for every point of increased NAEP difficulty (which corresponds roughly to one percent of the population, except in the extremes), about one percent (1.17 percent in grade 4 and 1.04 percent in grade 8) fewer students meet the standard. There is clearly much greater variability between states in the placement of mathematics standards than in the mathematics achievement of students: the standard deviations of state mean NAEP scale scores for the states included in this analysis are 7.0 points at grade 4 and 8.9 points at grade 8, compared to the standard deviations of their standards placement of 13.7 points and 16.7 points (table 3).

Figure 5. Relationship between the NAEP equivalents of grade 4 primary state mathematics standards and the percentages of students meeting those standards: 2003
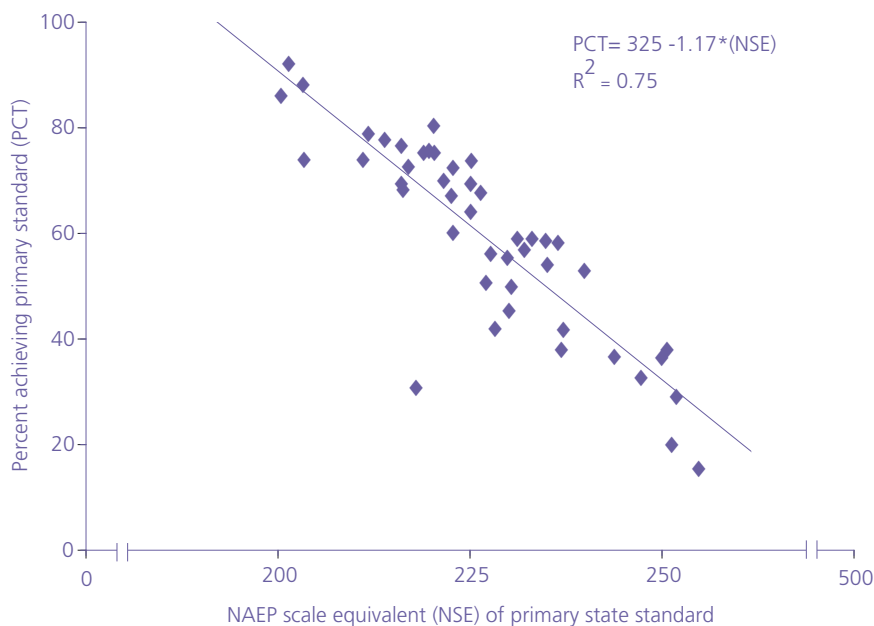
NOTE: Primary standard is the state's standard for *proficient* performance. Each diamond in the scatter plot represents the primary standard for one state. The relationship between the NAEP scale equivalent of grade 4 primary state mathematics standards (NSE) and the percentages of students meeting those standards in a state (PCT) is estimated over the range of data values by the equation PCT = 325 - 1.17(NSE). In other words, a one point increase in the NAEP difficulty implies 1.17 percent fewer students meeting the standard. For example, the 220 point on the NAEP scale equivalent represents approximately 67.6 percent of students achieving primary standard (67.6 = 325 - 1.17(220)) and at 221 on the same scale indicates 66.4 percent (325 - 1.17(221) = 66.4).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 6. Relationship between the NAEP equivalents of grade 8 primary state mathematics standards and the percentages of students meeting those standards: 2003
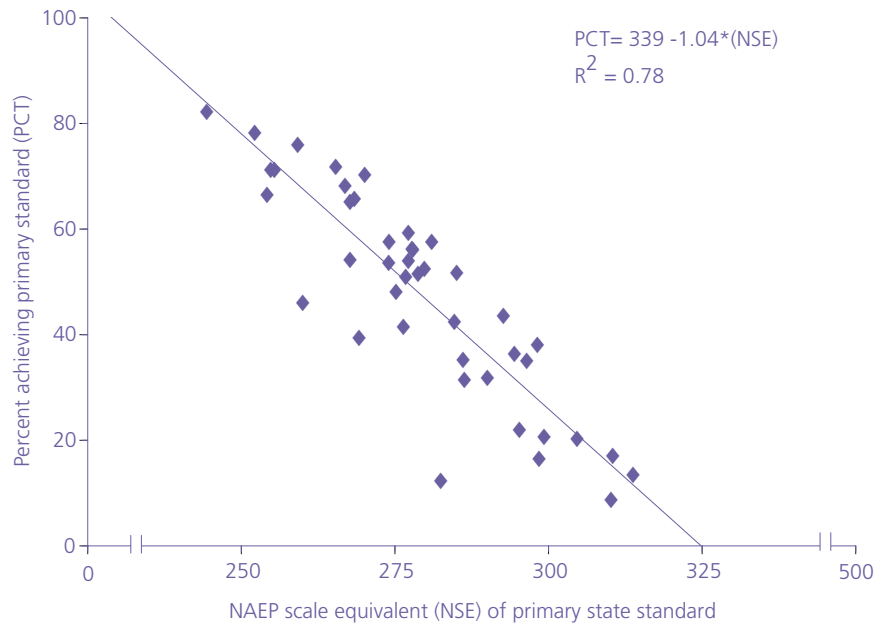
NOTE: Primary standard is the state's standard for *proficient* performance. Each diamond in the scatter plot represents the primary standard for one state. The relationship between the NAEP scale equivalent of grade 8 primary state mathematics standards (NSE) and the percentages of students meeting those standards (PCT) is estimated over the range of data values by the equation

PCT = 339 - 1.04(NSE). In other words, a one point increase in the NAEP difficulty implies 1.04 percent fewer students meeting the standard. For example, the 270 point on the NAEP scale equivalent represents approximately 58.2 percent of students achieving primary standard (58.2 = 339 - 1.04(270)) and at 271 on the same scale indicates 57.2 percent (339 - 1.04(271) = 57.2).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

How is variation among performance standards related to the performance of students on NAEP?

Does setting high standards lead to higher achievement? Finding out whether it does must await the accumulation of trend information over time, but the relation between the difficulty level of a state's primary mathematics standard and the performance of that state's students on the NAEP mathematics assessment is relevant. This question is addressed in figures 7 and 8, which display the percentage of each state's students meeting the NAEP proficient standard as a function of the placement of their own primary mathematics standard.

These graphs present a stark contrast to the relations shown in figures 5 and 6. In 2003, there was virtually no relation between the level at which a state sets its primary mathematics standard and the mathematics achievement of its students on NAEP. In most states, between 30 and 60 percent of students meet the NAEP proficient standard, and that percentage is no higher among states that set high primary standards than among states that set low primary standards.

Figure 7.   Relationship between the NAEP equivalents of grade 4 primary state mathematics standards and the percentages of students meeting the NAEP mathematics proficiency standard: 2003
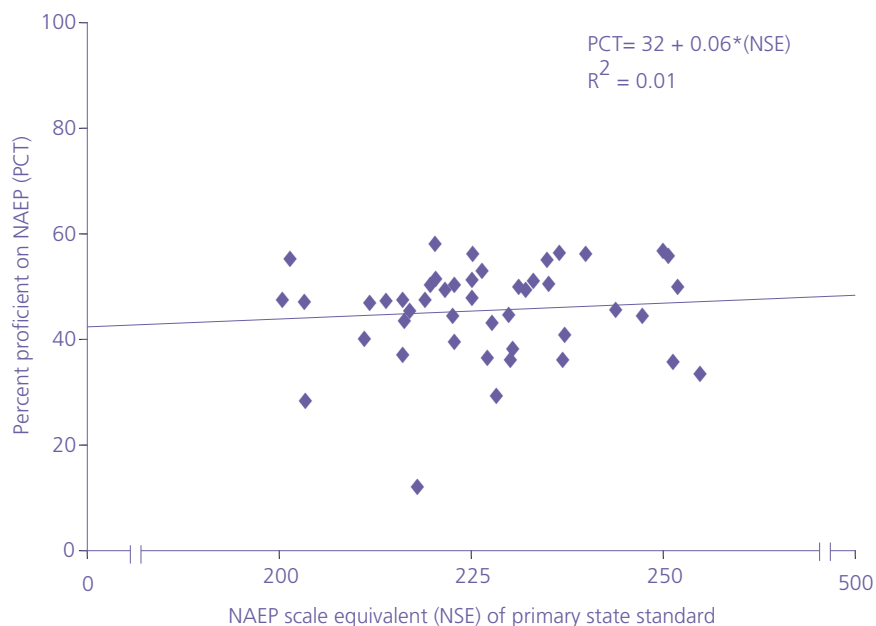
NOTE: Primary standard is the state's standard for *proficient* performance. The relationship between the NAEP scale equivalent of grade 4 primary state mathematics standards (NSE) and the percentages of students meeting NAEP mathematics proficiency standard (PCT) is estimated over the range of data values by the equation PCT = 32 + 0.06(NSE). There is virtually no relation between the level at which a state sets its primary mathematics standard and the mathematics achievement of its students on NAEP.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 8.   Relationship between the NAEP equivalents of grade 8 primary state mathematics standards and the percentages of students meeting the NAEP mathematics proficiency standard: 2003
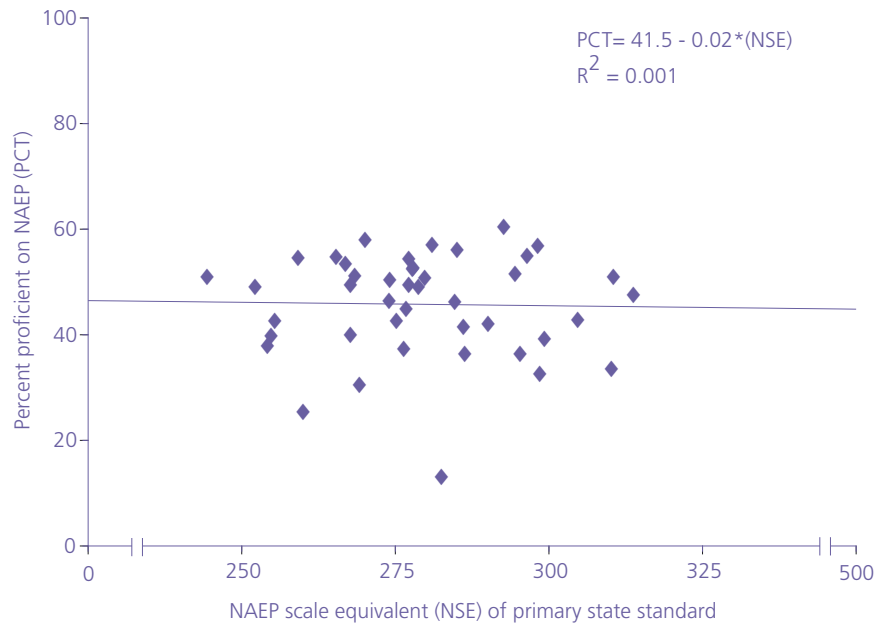
NOTE: Primary standard is the state's standard for *proficient* performance. The relationship between the NAEP scale equivalent of grade 8 primary state mathematics standards (NSE) and the percentages of students meeting NAEP mathematics proficiency standard (PCT) is estimated over the range of data values by the equation PCT = 41.5 - 0.02(NSE). There is virtually no relation between the level at which a state sets its primary mathematics standard and the mathematics achievement of its students on NAEP.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Is there a pattern in the placement of a state's performance standards relative to the range of student performance in the state?

As we saw in figures 5 and 6, the placement of the standards can have consequences for the ability to demonstrate school-level gains. It is therefore useful to see where states are setting their standards, single and multiple alike. The scatter plots in figures 9 and 10 extend the charts of primary standards shown in figures 3 and 4 to show the entire range of 130 grade 4 and 120 grade 8 state mathematics standards. In these scatter plots, standards higher than the primary standard are shown as plus/minus signs, primary standards as open/filled diamonds, and lower standards as open/filled circles. The 34 grade 4 standards and 5 grade 8 standards that have sufficiently high relative errors to question the validity of the mapping are indicated by dashes and

unfilled diamonds and circles, and are to the right of the vertical line in each figure. Grade 8 mathematics standards were more easily equated than grade 4 mathematics standards.

But how is this variability related to the student populations in the states? This question is addressed in an exploratory manner in figures 11 and 12, which display the frequencies of standards met by differing percentages of the population.[21] Thus, for example, the relatively easiest 14 standards for grade 4 were achieved by more than 90 percent of the students in their respective states, and the highest 11 standards were achieved by fewer than 10 percent of the students (figure 11).[22] At grade 8, five standards were achieved by more than 90 percent of the students in their respective states, while 14 were achieved by fewer than 10 percent (figure 12).

Standards for grade 4 (or grades 3 or 5, where there is no grade 4 state mathematics assessment) are set at every level, from very easy (more than 90 percent passing) to very difficult (fewer than 10 percent passing). The same is true for standards for grade 8 (or grade 7, where there is no grade 8 state mathematics assessment), although a greater percentage of the grade 8 standards are set to be difficult for eighth graders to pass: 38 of 120 are set where they are passed by fewer than 20 percent of eighth graders, compared to 28 of 130 grade 4 standards.[23] NAEP basic, proficient, and advanced mathematics achievement levels, by comparison, are met by about 70 percent (77 percent at grade 4 and 68 percent at grade 8), 30 percent, and 5 percent, respectively, of students nationally.

We conclude this section on state standards by pointing out the assumptions made in these comparisons. The major assumption is that the state assessment results are correlated with NAEP results—although the tests may look different, it is the correlation of their results that is important. If NAEP and the state assessment identify the same pattern of high and low achievement across schools in the state, then it is meaningful to identify NAEP scale equivalents of state assessment standards. The question of correlation is discussed in the next section.

The other important assumption is that the assessments are measuring the same population, in the same way. If substantial numbers of students participate in one of the assessments but not the other, this can have a biasing effect on the standard comparison. While we cannot account for state assessment non-participation in this comparison, we do account for NAEP non-participation by use of weighting and imputation of achievement of excluded students (see appendix A for a discussion of the imputation).

---

21. The grade 4 and grade 8 standards include a few that are for adjacent grades, as indicated in table 4, below.

22. If most students in a state can pass a performance standard, the standard must be considered relatively easy, even if fewer students in another state might be able to pass it.

23. On a simple test of proportions, the probability of this pattern is less than .05, but inferring statistical significance from this would be inappropriate because there was no *a priori* hypothesis of finding this pattern.

Figure 9.   NAEP equivalents of state grade 4 primary mathematics achievement standards, including standards higher and lower than the primary standards, by relative error criterion: 2003
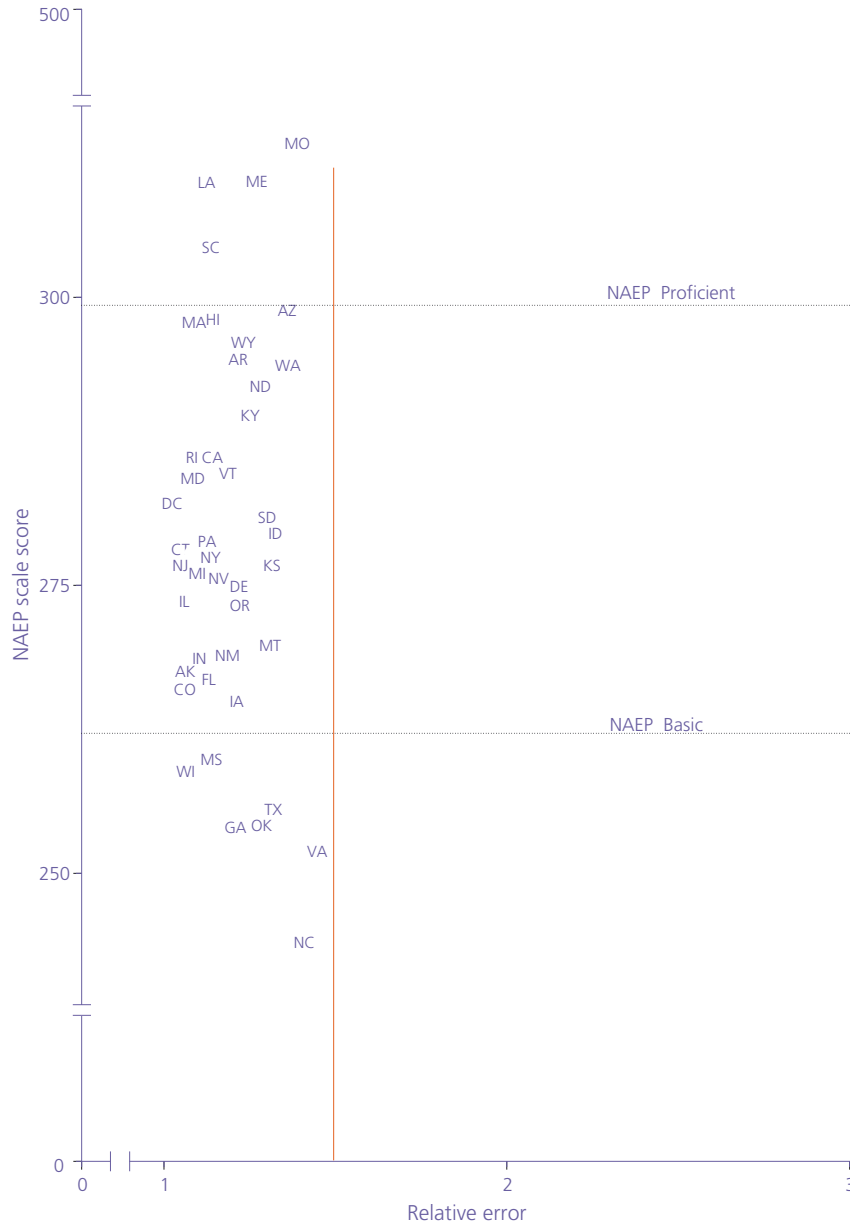

NOTE: Primary standard is the state's standard for *proficient* performance. Relative error is a ratio measure of reproducibility of school-level percentages meeting standards, described in appendix A. The vertical line indicates a criterion for maximum relative error.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 10. NAEP equivalents of state grade 8 primary mathematics achievement standards, including standards higher and lower than the primary standards, by relative error criterion: 2003

NOTE: Primary standard is the state's standard for *proficient* performance. Relative error is a ratio measure of reproducibility of school-level percentages meeting standards, described in appendix A. The vertical line indicates a criterion for maximum relative error.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 11. Number of state mathematics standards by percentages of grade 4 students meeting them: 2003
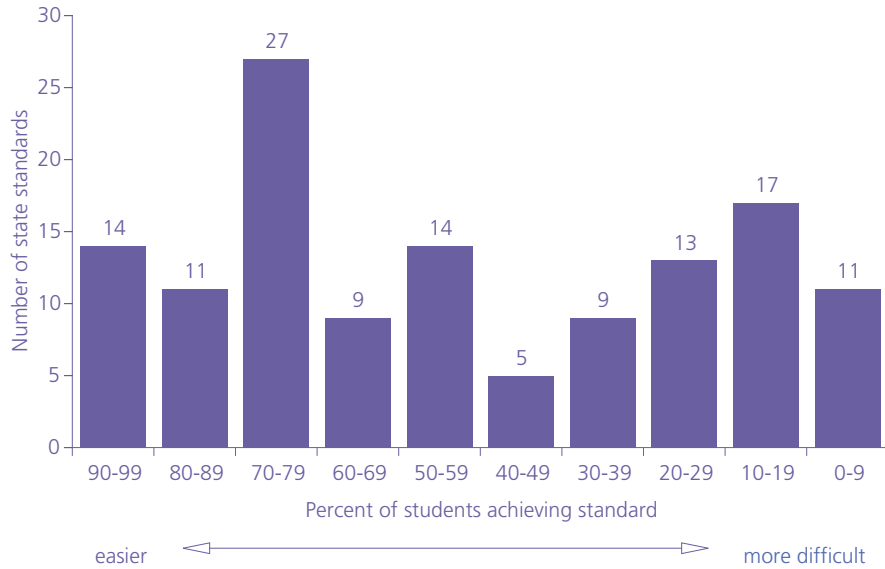
SOURCE: The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 12. Number of state mathematics standards by percentages of grade 8 students meeting them: 2003
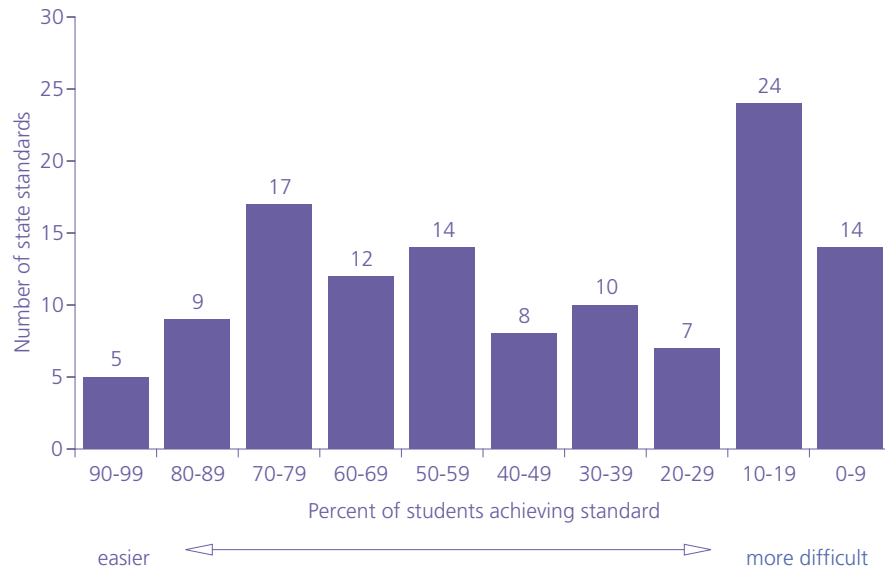
SOURCE: The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Finally, there is the issue of accommodations, or non-standard test administrations, provided for some students with disabilities and English language learners. It is not known at present how these accommodations (e.g., extended time and one-on-one testing) affect the distribution of assessment scores.

SUMMARY

By matching percentages of students reported to be meeting state standards in schools participating in NAEP with the distribution of performance of students in those schools on NAEP, cutpoints on the NAEP scale can be identified that are equivalent to the state standards. The accuracy of the determination of the NAEP equivalent of the standard depends on the correlations between NAEP and state assessment results. Most states have multiple standards, and these can be categorized into *primary standards*, which are generally the standards used for reporting adequate yearly progress, and standards that are above or below the primary standards. In most states, the primary standards are referred to as *proficient* or *meets the standard.*

In the majority of the states examined, the standards were sufficiently correlated to warrant reporting the NAEP equivalents of standards. At grade 4, 8 of the 46 primary standards were judged to be questionable; none were at grade 8. The mapping of state standards to the NAEP scale is an essential step in comparing achievement trends and gaps as measured by NAEP and state assessments.

The primary standards vary significantly in difficulty, as reflected in their NAEP equivalents. On average, for any two randomly selected states, about 15 percent of the students who meet the primary standard in one state would not meet the standard in the other state; between some states, the disparity is much larger.

As might be expected, the higher the primary standard is set, the fewer the students who meet it. Between states, there is more variability in the setting of standards than in actual student achievement. On the other hand, students in states with high primary standards score just about the same on NAEP as students in states with low primary standards.

3 Comparing Schools' Percentages Meeting State Standards

A fundamental question is whether state assessments in different states would identify the same schools as having high and low mathematics achievement. No state assessments are administered exactly the same way, with the same content, across state lines, so analysts cannot answer this question directly. However, NAEP provides a link, however imperfect, to address this question. If the pattern of NAEP results matches the pattern of state assessment results across schools in each of two different states, then either of those two states' assessments would likely identify the same schools as having students who are good in mathematics, compared to other schools in their respective states.

The correlation coefficient is the standard measure of the tendency of two measurements to give the same results, varying from +1 when two measurements give functionally identical results, to 0 when they are completely unrelated to each other, to –1 when they represent opposites. A high correlation (near +1) between two mathematics achievement tests means that schools (or students) whose performance is relatively high on one test also demonstrate relatively high performance on the other test. It does not mean that the two tests are necessarily similar in content and format, only that their results are similar. And it is the results of the tests that are of concern for accountability purposes.

To compute a correlation coefficient, one needs the results of both tests for the same schools (or students). State assessment statistics are available at the school level, and NAEP data can be aggregated from student records to create school-level statistics for the same schools.[24] Therefore, the correlations presented in this report are of school-level statistics, and a high correlation indicates that two assessments are identifying the same schools as high scoring (and the same schools as low scoring).[25]

---

24. NAEP does not report individual school-level statistics because the design of NAEP precludes measurement of school-level statistics with sufficient accuracy (reliability) to justify public release. However, for analytical purposes, aggregating these school-level statistics to state-level summaries provides reliable state-level results (e.g., correlations between NAEP and state assessment results).

25. The value of a correlation coefficient at the student level and one at the school level need not be the same, even though they are based on the same basic data. The student level correlation will tend to be somewhat lower because it does not give as much weight to systematic variation in distributions of higher and lower achieving students to different schools. Because policy analysts are interested in systematic variation between schools, the school-level correlation is the appropriate statistic.

State assessments have traditionally reported scores in a wide variety of units, including percentile ranks, scale scores, and grade equivalent scores, among others, but since 1990 there has been a convergence on reporting in terms of the percentages of students meeting standards (which is translated into the percentages of students earning a score above some *cutpoint*). While this does not present an insurmountable problem for computation of correlation coefficients, it does raise three issues that need to be addressed.

Most important of these is the match between the two standards being correlated. The correlation between the percentage of students achieving a very easy standard (e.g., one which 90 percent of students pass), with the percentage of students achieving a very hard standard (e.g., one which only 10 percent of students pass) will necessarily be lower than the correlation between two standards of matching difficulty. For this reason, the correlations presented in this report are between (a) the school-level percentages meeting a state's standards as measured by its own measurement, and (b) the corresponding percentages meeting a standard of the same difficulty as measured by NAEP. In the preceding chapter, NAEP cutpoints of difficulty equivalent to state standards were identified (in figures 3 and 4), and they are used in this analysis.

The second issue concerns the position of the standards in the achievement distribution even when they are matched in difficulty. Extremely easy or difficult standards necessarily have lower intercorrelations than standards near the median of the population. It is impossible to dictate where a state's standards fall in its achievement distribution, but it is possible to estimate how much the extremity of the standards might affect correlations.

The third issue concerns the fact that percentages meeting standards necessarily hide information about variations in achievement within the subgroup of students who meet the standard (and within the subgroup of students who fail to meet the standard). One might expect this to set limits on the correlation coefficients. However, empirical comparison of correlations of percentages meeting standards near the center of the distribution with correlations of median percentile ranks or mean scale scores has indicated that there is only modest loss of correlational information in using percentages meeting standards near the center of the distribution (MacLaughlin, 2005 and Shkolnik and Blankenship, 2006).

The correlations between the percentage of schools' students meeting the NAEP and the state assessment primary standards are shown in table 4. The selection of the standard and the short name of the standard included in the table are based on interpretation of information on the state's web site. The grade indicated is generally the same as NAEP (4 or 8), but in a few states, scores were available only for grade 3, 5, 7, or 9 (or for E or M, which represent aggregates across elementary or middle grades). The correlations for primary standards range from .44 to .89, with a median of .76, for grade 4 mathematics, and from .62 to .97, with a median of .81, for grade 8 mathematics. The distributions of correlations are shown in figures 13 and 14.

Table 4.   Correlations between NAEP and state assessment school-level percentages meeting primary state mathematics standards, grades 4 and 8, by state: 2003

State/ jurisdiction	Name of standard	Grades for state assessment		Grade 4 correlation	Grade 8 correlation
Alabama	Percentile Rank	4	8	0.80	0.84
Alaska	Proficient	4	8	0.78	0.86
Arizona	Meeting	5	8	0.77	0.69
Arkansas	Proficient	4	8	0.81	0.77
California	Proficient	4	7	0.84	0.88
Colorado	Partially Proficient	5	8	0.79	0.87
Connecticut	Goal	4	8	0.89	0.89
Delaware	Meeting	5	8	0.58	0.79
District of Columbia	Proficient	4	8	0.69	0.97
Florida	(3)PartialSuccess	4	8	0.89	0.86
Georgia	Meeting	4	8	0.83	0.80
Hawaii	Meeting	5	8	0.78	0.83
Idaho	Proficient	4	8	0.67	0.70
Illinois	Meeting	5	8	0.84	0.92
Indiana	Pass	3	8	0.44	0.83
Iowa	Proficient	4	8	0.77	0.77
Kansas	Proficient	4	7	0.66	0.72
Kentucky	Proficient	5	8	0.53	0.72
Louisiana	Mastery	4	8	0.79	0.82
Maine	Meeting	4	8	0.56	0.69
Maryland	Proficient	5	8	0.83	0.88
Massachusetts	Proficient	4	8	0.82	0.87
Michigan	Meeting	4	8	0.74	0.87
Minnesota	(3)Proficient	5	—	0.77	—
Mississippi	Proficient	4	8	0.79	0.82
Missouri	Proficient	4	8	0.69	0.62
Montana	Proficient	4	8	0.72	0.72
Nebraska	Meeting	—	—	—	—
Nevada	Meeting:3	4	7	0.81	0.82
New Hampshire	Basic	3	—	0.46	—
New Jersey	Proficient	4	8	0.84	0.90
New Mexico	Top half	4	8	0.77	0.81
New York	Meeting	4	8	0.86	0.85
North Carolina	Consistent Mastery	4	8	0.63	0.71
North Dakota	Meeting	4	8	0.64	0.75
Ohio	Proficient	4	—	0.81	—
Oklahoma	Satisfactory	5	8	0.58	0.71
Oregon	Meeting	5	8	0.51	0.77
Pennsylvania	Proficient	5	8	0.83	0.87
Rhode Island	Proficient	4	8	0.78	0.90
South Carolina	Proficient	4	8	0.74	0.80
South Dakota	Proficient	4	8	0.77	0.71
Tennessee	Percentile Rank	4	8	0.76	0.81
Texas	Passing	4	8	0.52	0.71
Utah	Percentile Rank	5	8	0.68	0.72
Vermont	Achieved	4	8	0.47	0.74
Virginia	Proficient	5	8	0.54	0.63
Washington	Met	4	7	0.69	0.69
West Virginia	Top half	—	—	—	—
Wisconsin	Proficient	4	8	0.81	0.90
Wyoming	Proficient	4	8	0.64	0.74

— Not available.

NOTE: Primary standard is the state's standard for *proficient* performance. In Alabama, Tennessee, and Utah, correlations are based on school-level median percentile ranks. In West Virginia, E and M represent aggregates across elementary and middle grades, respectively.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 13. Frequency of correlations between school-level NAEP and state assessment percentages meeting the primary grade 4 state mathematics standard: 2003

NOTE: Primary standard is the state's standard for *proficient* performance. No correlations lie on the boundaries of the categories. Correlations are of median percentile ranks for Alabama, Tennessee, and Utah.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 14. Frequency of correlations between school-level NAEP and state assessment percentages meeting the primary grade 8 state mathematics standard: 2003

NOTE: Primary standard is the state's standard for *proficient* performance. No correlations lie on the boundaries of the categories. Correlations are of median percentile ranks for Alabama, Tennessee, and Utah.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

As an overall criterion, one would like to have correlations greater than .70 for analyses that depend on a linkage between the results of two assessments. In states that do not meet this criterion, i.e., the two assessments have less than 50 percent of common variance, divergences in comparisons of trends and gaps in later sections of this report may reflect the impact of whatever factors cause the correlations to be lower. Because this is a research report, we do not exclude data from states with lower correlations.

There are many sources of variation in correlation coefficients; results presented here can only set a context for in-depth analysis of the differences which analysts may wish to pursue. The tendency to say that "they must be measuring different things" should be resisted, however. Even if the tests were sampling and measuring different parts of the mathematics construct, they still might be highly correlated; that is, they might still identify the same schools as high achieving and low achieving.

The following (non-exhaustive) list of reasons for lower correlations should be considered before selecting any particular interpretation of low correlations.

- Reliability of the measure (the school-level test score)
  - Student sample size in schools (small school suppression may increase correlation)
  - Reliability of the student-level measure
  - Measures from different grades
- Conditions of testing
  - Different dates of testing (including testing in different grades)
  - Different motivation to perform
- Requirements for enabling skills
  - Different response formats (different demands for writing skills)
- Similarity of location of the measure relative to the student population
  - Extreme standards will not be as strongly correlated as those near the median
- Similarity of testing accommodations provided for students with special needs
  - Accommodations given on one test but not the other reduce correlations
- Match of the student populations included in the statistics
  - Representativeness of NAEP samples of students and of schools
  - Extent of student exclusion or non-participation
- Differences in the definition of the target skill (mathematics)
  - States have varying emphasis on computation, problem-solving, and conceptual understanding.

To understand the potential impact of these factors, consider the effects of three factors on the correlations of NAEP and state assessment percentages meeting standards: (1) extremity of the standard, (2) size of the school sample of students on which the percentage is based, and (3) grade level of testing (same grade or adjacent grade). As a meta-analysis of the correlation coefficients, we carried out a linear

regression accounting for variation in correlations for 130 standards in 46 states in grade 4 and 120 standards in 43 states in grade 8.[26] Results are shown in table 5.

Table 5.   Standardized regression coefficients of selected factors accounting for variation in NAEP-state mathematics assessment correlations: 2003

Factor	Grade 4	Grade 8
Extreme standards	-0.51 *	-0.54 *
Small school samples	-0.44 *	-0.20 *
Grade difference	-0.27 *	-0.11
Sample size	130	120
R^2	0.47	0.32

* Coefficient statistically significantly different from zero ($p<.05$)

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

The values of R^2 were .47 and .32 for grades 4 and 8. That is, these three factors accounted for one-half of the variance of correlations between NAEP and state standards in grade 4 and one-third of the variance in grade 8 correlations. At grade 4, all three predictors were significant, but for grade 8, the effect of the grade difference factor was not significantly different from zero. This may be due to the fact that in only four states did we have to use grade 7 state assessment scores.

Applying the results of the linear regression, one can estimate what each correlation might have been if it were based on a standard set at the student population median, in the same grade as NAEP, and with no school samples of fewer than 30 students. The results are displayed in table 6 and summarized in figures 15 and 16. Nearly all of the adjusted correlations are greater than .70 for grade 8, with a median of .86, and four-fifths of them are at least .70 for grade 4, with a median of .84.

At grade 8, correlations in two states remained less than .70, after adjusting for effects of grade differences, small schools, and extreme standards, although their correlations rounded to .68 (Virginia) and .70 (Idaho). At grade 4, adjusted correlations in nine states were less than .70: Delaware, Indiana, Kentucky, New Hampshire, Oregon, Virginia, Texas, Vermont, and Washington. In the first six of these states, the correlations were based on state assessment scores for grade 3 or 5. It is possible that the adjustment did not capture all of the effects of that factor. The factors affecting the correlation coefficients in the other three states are not known at this time.

---

26. All state standards, not merely the primary one for each state, were included. The specific predictors were: (1) $(d/50)^4$, where d was the difference between the average percentage meeting the standard and 50 percent; (2) the maximum of 0 and the amount by which the average NAEP school's student sample size was less than 30 (34 was the largest average school sample size in grade 4); and (3) a dichotomy, 1 if the tested grade was not 4 or 8, 0 if it was 4 or 8.

Table 6.   Adjusted correlations between NAEP and state assessment school-level percentages meeting primary state mathematics standards, grades 4 and 8, by state: 2003

State/ jurisdiction	Name of standard	Grades for state assessment		Grade 4 correlation	Grade 8 correlation
Alabama	Percentile Rank	4	8	—	—
Alaska	Proficient	4	8	0.98	0.90
Arizona	Meeting	5	8	0.87	0.79
Arkansas	Proficient	4	8	0.84	0.86
California	Proficient	4	7	0.84	0.94
Colorado	Partially Proficient	5	8	1.00	0.93
Connecticut	Goal	4	8	0.89	0.91
Delaware	Meeting	5	8	0.69	0.79
District of Columbia	Proficient	4	8	0.79	1.00
Florida	(3)Partial Success	4	8	0.89	0.89
Georgia	Meeting	4	8	0.84	0.81
Hawaii	Meeting	5	8	0.93	0.92
Idaho	Proficient	4	8	0.73	0.70
Illinois	Meeting	5	8	0.93	0.96
Indiana	Pass	3	8	0.55	0.85
Iowa	Proficient	4	8	0.88	0.82
Kansas	Proficient	4	7	0.80	0.83
Kentucky	Proficient	5	8	0.64	0.77
Louisiana	Mastery	4	8	0.91	1.00
Maine	Meeting	4	8	0.75	0.77
Maryland	Proficient	5	8	0.92	0.91
Massachusetts	Proficient	4	8	0.86	0.88
Michigan	Meeting	4	8	0.76	0.92
Minnesota	(3)Proficient	5	—	0.86	—
Mississippi	Proficient	4	8	0.80	0.86
Missouri	Proficient	4	8	0.71	0.76
Montana	Proficient	4	8	0.97	0.82
Nebraska	Meeting	—	—	—	—
Nevada	Meeting:3	4	7	0.81	0.86
New Hampshire	Proficient	3	—	0.65	—
New Jersey	Proficient	4	8	0.84	0.93
New Mexico	Top half	4	8	0.85	0.81
New York	Meeting	4	8	0.89	0.90
North Carolina	Consistent Mastery	4	8	0.80	0.76
North Dakota	Meeting	4	8	0.90	0.86
Ohio	Proficient	4	—	0.81	—
Oklahoma	Satisfactory	5	8	0.78	0.79
Oregon	Meeting	5	8	0.68	0.81
Pennsylvania	Proficient	5	8	0.93	0.89
Rhode Island	Proficient	4	8	0.79	0.90
South Carolina	Proficient	4	8	0.74	0.87
South Dakota	Proficient	4	8	0.98	0.79
Tennessee	Percentile Rank	4	8	—	—
Texas	Passing	4	8	0.60	0.71
Utah	Percentile Rank	5	8	—	—
Vermont	Achieved	4	8	0.69	0.78
Virginia	Proficient	5	8	0.66	0.68
Washington	Met	4	7	0.69	0.77
West Virginia	Top half	—	—	—	—
Wisconsin	Proficient	4	8	0.90	0.96
Wyoming	Proficient	4	8	0.87	0.75

— Not available.

NOTE: Primary standard is the state's standard for *proficient* performance. For Alabama, Tennessee, and Utah, adjusted correlations could not be estimated since percentages meeting standards were not available.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 15. Frequency of adjusted correlations between school-level NAEP and state assessment percentages meeting the primary grade 4 state mathematics standard: 2003

NOTE: Primary standard is the state's standard for *proficient* performance. No correlations lie on the boundaries of the categories. Correlations of median percentile ranks for Alabama, Tennessee, and Utah are not included.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 16. Frequency of adjusted correlations between school-level NAEP and state assessment percentages meeting the primary grade 8 state mathematics standard: 2003

NOTE: Primary standard is the state's standard for *proficient* performance. No correlations lie on the boundaries of the categories. Correlations of median percentile ranks for Alabama, Tennessee, and Utah are not included.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

The high adjusted correlations between NAEP and state assessment measures of the percentages of schools' students meeting mathematics achievement standards indicate that in most states, NAEP and state assessments in 2003 were in general agreement about which schools have high and low mathematics achievement. Nevertheless, the findings of relatively low correlations in a few states need to be considered in interpreting results of gap and trend comparisons of trends and gaps as reported by NAEP and state assessments. Gaps and trends may be similar, in spite of low correlations, but when gaps or trends differ significantly, the reasons for the low correlations require further study.

SUMMARY

An essential criterion for the comparison of NAEP and state assessment results in a state is that the two assessments agree on which schools are high achieving and which are not. The critical statistic for testing this criterion is the correlation between schools' percentages achieving the standard, as measured by NAEP and the state assessment.

In 2003, correlations between NAEP and state assessment measures of mathematics achievement were greater than .70 in 30 of 49 states for grade 4 and 41 of 46 states for grade 8. An analysis of the correlations focused on two methodological factors that tend to depress some of these correlations: (1) small enrollments in schools limit the reliability of percentages of students meeting a standard; and (2) standards set either very high or very low tend to be less correlated than standards set near the middle of a state's achievement distribution. Estimates of what the correlations would be if they were all based on scores on non-extreme standards in the same grade in schools with more than 30 students per grade resulted in correlations greater than .70 in 37 of 46 states/jurisdiction for grade 4 and 42 of 43 states/jurisdiction for grade 8.

4 Comparing Changes in Achievement

A central concern to both state and NAEP assessment programs is an examination of achievement trends over time (e.g., USDE 2002, NAGB 2001). The extent to which NAEP measures of achievement trends match states' measures of achievement trends may be of interest to state assessment programs, the federal government, and the public in general. The purpose of this section is to make direct comparisons between NAEP and state assessment changes over time.

Unlike state assessments, NAEP is not administered every year, and NAEP is only administered to a sample of students in a sample of schools in each state. NAEP sample schools also vary from year to year. For this comparison report, our comparison of changes in mathematics achievement is limited to those between the 1999-2000 and 2002-2003 school years (i.e., between 2000 and 2003). For research purposes, analysts may wish to examine trends in earlier NAEP years (e.g., 1991-1992 and 1995-1996), but the NLSLSASD does not have sufficient state assessment data for those early years to warrant inclusion in this report.

To make meaningful comparisons of gains between NAEP and the state assessments, we included only the NAEP sample schools for which state assessment scores were available in this trend analysis.[27] This allows us to eliminate effects of random or systematic variation between schools in comparing NAEP and state assessments.

There are many states for which we did not have scores in multiple years and so could not measure achievement changes over time. In addition to these states, there are others for which we could not use 2003 scores for the trend analysis because they changed their state assessments and/or primary standards in 2003; changes in percentages meeting the primary standards from 2000 to 2003 will not reflect their actual changes in achievement. Therefore, these states are excluded from the analysis; they are listed below along with the reasons for exclusion:

---

27. These schools were weighted, according to NAEP weights, to represent the state. The assumption being made by using these unadjusted NAEP weights is that any NAEP school without state scores would have state scores averaging close to the state mean. That can be tested by comparing the NAEP means for the schools with and without state scores in each state. It is our belief that there is so few of these schools that it doesn't matter since we are matching close to 100 percent of the schools. Moreover, the analyses such as standard estimation are based on the subset of schools with both scores, so they are not biased by the omission of a few NAEP schools.

- California: changed assessment in 2003
- Indiana: changed assessment in 2003
- Maryland: changed assessment in 2003
- Michigan: changed assessment in 2003
- Nevada: changed assessment in 2003
- Texas: changed assessment in 2003
- Virginia: changes performance standards every year and no longitudinal equating from year to year done.
- Wisconsin: set new performance standards in 2003

It is important to note that changes in percentage meeting the primary standards may be affected by ceiling effects. In other words, if a state sets a relatively low standard and many schools in the state show very high percentages of students already meeting the standard in the base year, there will be little "room to grow" for these schools. The state would be less likely to show positive achievement trends, not because students are not learning, but because many students have already met the standard in the base year.

Finally, all significance tests are of differences between NAEP and state assessment results. The comparisons between NAEP and state assessment results in each state are based on that state's primary state standard. This means that the standard at which the comparison is made is different in each state. For this reason, comparisons between states are not appropriate.

In table 7, we summarize the average of, and variation in, changes in achievement over time on NAEP and the state assessments across states in terms of percentages achieving the state primary standards.[28] The state primary standard is, in most cases, the standard used for reporting adequate yearly progress, in compliance with *No Child Left Behind.* We *rescored* NAEP in terms of percentages meeting the state primary standards, because comparisons of trends at differing locations in the distribution of achievement are not easily interpretable. To rescore NAEP for trend comparisons, we estimated the location of the state primary standard on the NAEP scale in the initial trend year. In that year, the NAEP and the state assessment percentages match by definition. Finally, the State Profile section of this report (Appendix D) compares trends for multiple state standards, not only the single primary standard.

---

28. For Alabama, Tennessee, and Utah, we did not have percentages meeting a state standard; instead, for these states we report trends in percentile ranks.

Table 7.   Mathematics achievement gains in percentage meeting the primary state standard in grades 4 and 8: 2000 and 2003

Statistic	Grade 4 2000 to 2003 (24 states)		Grade 8 2000 to 2003 (22 states)	
	State	**NAEP**	**State**	**NAEP**
Average gain	7.0	9.9 *	3.0	3.6
standard error	1.89	2.40	1.43	2.06
Between-state standard deviation	9.12	5.27 *	4.98	3.21 *
standard error	0.38	0.45	0.37	0.44

* NAEP gains are significantly different from gains reported by the state assessment ($p<.05$).

NOTE: Primary standard is the state's standard for *proficient* performance. State assessment gains are recorded here for the schools that participated in NAEP. Gains are weighted to represent the population in each state. Averages are based on states with scores on the same tests in the two years.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Averaged over the states in which gains from 2000 to 2003 could be compared, mathematics achievement gains reported by NAEP are larger than those in state assessment in both grades 4 and 8, although the differences are not statistically significant at grade 8. These results are reversed in reading: that is, reading achievement gains reported by state assessments are larger than those in NAEP in both grades 4 and 8 (McLaughlin et al., 2008).

There are many possible explanations for the differences in gains measured by NAEP and state assessments from 2000 to 2003 that are associated with differences in testing. We have constructed the comparisons to remove two important sources of error, by comparing trends for the same sets of schools and at the same standard level. Other factors to be considered include (1) differences in changes in accommodations provided on the two assessments; (2) differences in changes in the student populations in each sampled school (e.g., NAEP vs. all fourth graders); (3) differences in changes in motivation (low stakes/high stakes); (4) differences in test modality (e.g., multiple choice vs. constructed response); (5) differences in time of year; and (6) a recalibration of the state assessment between trend years (of which we are not aware).

In addition to differences in gains measured by NAEP and state assessments, variations in gains across states are also of interest. As shown in table 7, the gains in percentages meeting the primary standards measured by state assessment vary substantially between states. The standard deviations of these gains vary from five to nine percentage points. However, these differences may overestimate the actual variation in gains in different states. The standard deviation of gains between states is smaller when they are measured by a common assessment, NAEP.

Interpreting these variations requires caution. Gains are measured at different points on the achievement continuum in different states; therefore, the gains are not comparable across states. However, we believe that the search for common trends

across states provides us with valuable information to make descriptive statements about general patterns in the nation.

It is possible that the greater variation in state assessment gains than in NAEP gains is due to different states' measurement of unique aspects of mathematics achievement not fully addressed by NAEP. However, an alternative hypothesis must be considered before searching for the unique aspects of mathematics achievement measured in states with relatively large gains: that is, a substantial portion of the variation in state assessment gains may be due to methodological differences in the way that state assessments measure gains.

If the NAEP and state assessment gains are correlated positively, both assessments are likely to identify the same schools as making achievement gains. In other words, a school identified as increasing achievement by the state assessment is likely to be identified as increasing achievement by NAEP as well. To investigate whether the NAEP gains and state assessment gains are related, we present below scatter plots between the NAEP and state assessment gains for both grades 4 and 8.

Figure 17 indicates that the relationship between the NAEP and state assessment gains in grade 4 mathematics is positive, with R^2 of .28 and the correlation coefficient of .52 (with p-value<.05). This means that, in grade 4 mathematics, states that increased the percentage meeting the primary standard in the state assessment tend to be the states that increased the corresponding percentage in NAEP. Grade 4 mathematics is the only grade and subject that showed a relationship between the NAEP and state assessment gains: the results for grade 8, displayed in figure 18, and the reading results (McLaughlin et al. 2008) suggest that states in which the percentage of students meeting the primary reading standard based on the state assessment increased are not necessarily the states in which the percentage meeting the primary standard on NAEP increased.

Since the state assessment results vary significantly from state to state, it is important to identify particular states where the NAEP and the state mathematics assessment trends differ. A search for explanations of the different results must begin with identification of states in which they differ. State-by-state comparisons between NAEP and state assessment measurements of mathematics achievement trends are presented in the following tables.

Figure 17. Relationship between NAEP and state assessment gains in percentage meeting the primary state grade 4 mathematics standard: 2000 to 2003
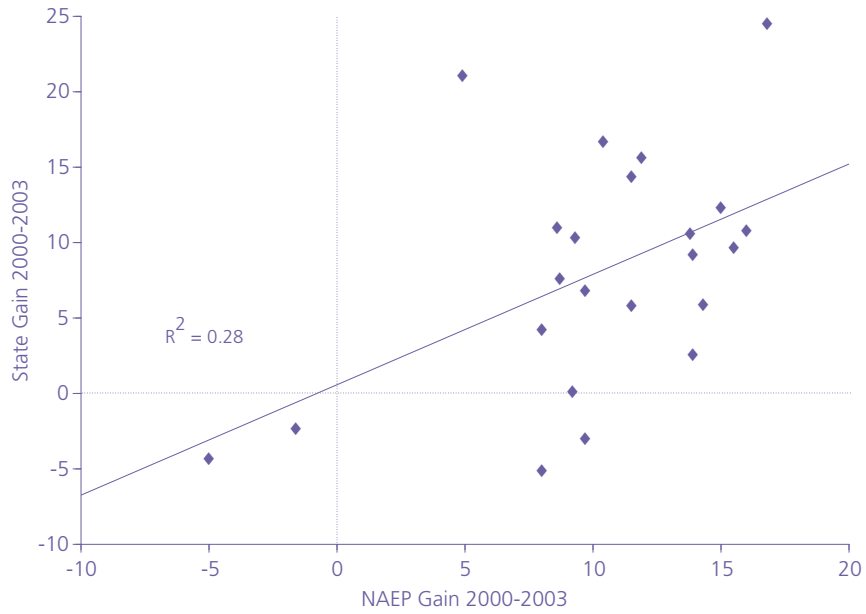
NOTE: Primary standard is the state's standard for *proficient* performance.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 18. Relationship between NAEP and state assessment gains in percentage meeting the primary state grade 8 mathematics standard: 2000 to 2003
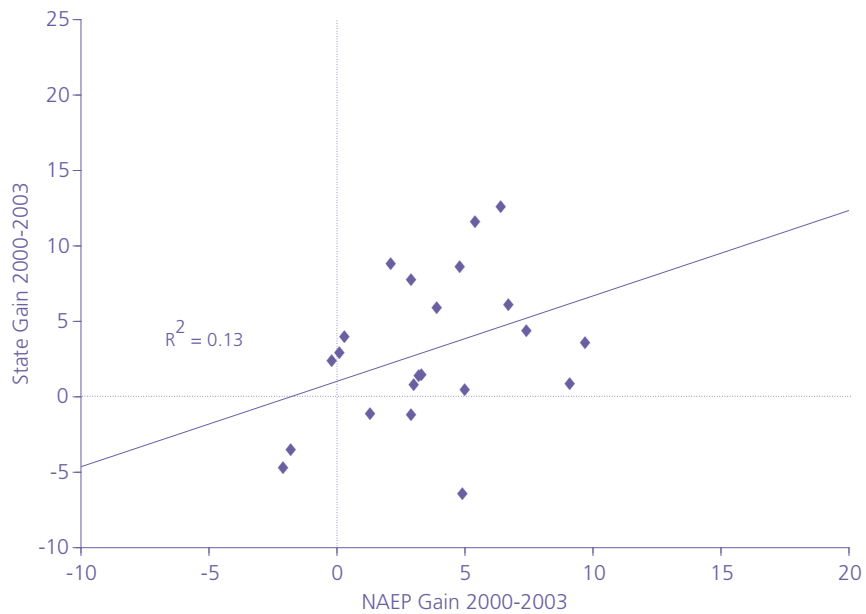
NOTE: Primary standard is the state's standard for *proficient* performance.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table 8 summarizes state-by-state trends for grade 4 mathematics achievement. Eleven out of 24 states increased percentages meeting the primary standard in NAEP significantly more than in their state assessments from 2000 to 2003; these are Arizona, Connecticut, Hawaii, Louisiana, Maine, Massachusetts, Missouri, Oklahoma, South Carolina, Vermont, and Wyoming. On the other hand, in Arkansas, Ohio, and Rhode Island, state assessments found significantly greater gains from 2000 to 2003 in mathematics than NAEP did.

Table 9 summarizes state-by-state trends for grade 8 mathematics scores. Between 2000 and 2003, NAEP found significantly larger mathematics achievement gains than their state assessments in six out of 22 states; these are Connecticut, Hawaii, Massachusetts, Missouri, South Carolina, and Wyoming. All six states observed the same trends in their grade 4 mathematics results: that is, significantly greater gains in NAEP than in their state assessments. On the other hand, in Arkansas, Georgia, Illinois, New York, and Rhode Island, the state assessments measured significantly greater gains from 2000 to 2003 in the percentage meeting the state primary standard than NAEP did.

Do the states in which we found significant discrepancies between NAEP and state assessment trends differ from the other states? One reasonable explanation for trend differences is that the state assessments do not identify the same schools as high and low achieving that NAEP does. Since we *rescored* NAEP in terms of percentages meeting the state primary standards so that the NAEP and the state assessment percentages match in the base year (i.e., 2000), we address this issue by comparing a) the correlation between 2003 NAEP and state assessment results (table 4) in the states with significant discrepancies (in tables 8 and 9) to b) the correlation in other states without such discrepancies.

Overall, there is no noticeable difference between the states with significant discrepancies and those without such discrepancies.[29] In the grade 4 results, the average correlation between NAEP and state assessments in 2003 is .72 for the states with significant discrepancies in gains and .71 for other states without such discrepancies. Patterns are very similar in grade 8. The average correlation between NAEP and state assessments is .82 for the states with significant discrepancies in gains from 2000 to 2003 and .74 for other states without discrepancies. These results indicate that in both grades 4 and 8, there is no relationship between: a) the tendency for the two assessments to identify the same schools as low achieving and high achieving in 2003, and b) the sizes of discrepancies in gains as measured by NAEP and by state assessments.

---

29. The statement is based on the fact that the correlations are similar. No statistical tests were performed on the differences.

National Assessment of Educational Progress

Table 8.   Mathematics achievement gains in percentage meeting the primary standard in grade 4, by state: 2000 and 2003

State/ jurisdiction	State			NAEP		
	2000	2003	Gain	2000	2003	Gain
Alabama	57.4	53.1	-4.3	41.7	36.7	-5.0
Alaska	—	67.2	—	—	67.2	—
Arizona	40.6	49.8	9.2	40.6	54.5	13.9 *
Arkansas	35.6	60.1	24.5	35.7	52.5	16.8 *
California	30.9	—	—	30.9	—	—
Colorado	—	86.0	—	—	86.0	—
Connecticut	63.7	58.6	-5.1	63.7	71.7	8.0 *
Delaware	—	72.7	—	—	72.6	—
District of Columbia	36.9	—	—	36.9	—	—
Florida	—	55.4	—	—	55.4	—
Georgia	61.6	73.9	12.3	61.7	76.7	15.0
Hawaii	64.8	67.4	2.6	64.8	78.7	13.9 *
Idaho	—	76.6	—	—	76.7	—
Illinois	52.7	68.3	15.6	52.8	64.7	11.9
Indiana	—	—	—	—	—	—
Iowa	—	75.2	—	—	75.2	—
Kansas	59.3	73.7	14.4	59.3	70.8	11.5
Kentucky	31.1	37.9	6.8	31.1	40.8	9.7
Louisiana	11.3	15.5	4.2	11.4	19.4	8.0 *
Maine	23.3	29.1	5.8	23.3	34.8	11.5 *
Maryland	45.6	—	—	45.6	—	—
Massachusetts	41.0	38.0	-3.0	40.9	50.6	9.7 *
Michigan	76.9	—	—	77.1	—	—
Minnesota	47.9	58.2	10.3	47.9	57.2	9.3
Mississippi	—	74.0	—	—	73.9	—
Missouri	36.6	36.7	0.1	36.6	45.8	9.2 *
Montana	—	75.3	—	—	75.4	—
Nebraska	60.0	—	—	60.0	—	—
Nevada	—	—	—	—	—	—
New Hampshire	—	80.3	—	—	80.3	—
New Jersey	—	67.6	—	—	67.5	—
New Mexico	—	42.0	—	—	42.0	—
New York	67.8	78.8	11.0	67.8	76.4	8.6
North Carolina	84.6	92.2	7.6	84.5	93.2	8.7
North Dakota	—	59.0	—	—	59.0	—
Ohio	42.3	59.0	16.7	42.5	52.9	10.4 *
Oklahoma	85.8	69.3	-16.5	85.9	90.3	4.4 *
Oregon	67.2	77.8	10.6	67.2	81.0	13.8
Pennsylvania	—	56.8	—	—	56.8	—
Rhode Island	20.7	41.8	21.1	20.7	25.6	4.9 *
South Carolina	22.9	32.6	9.7	23.0	38.5	15.5 *
South Dakota	—	72.5	—	—	72.5	—
Tennessee	—	54.5	—	—	41.9	—
Texas	88.5	—	—	88.5	—	—
Utah	50.2	47.9	-2.3	52.9	51.3	-1.6
Vermont	47.0	52.9	5.9	46.9	61.2	14.3 *
Virginia	—	—	—	—	—	—
Washington	—	54.0	—	—	54.0	—
West Virginia	—	—	—	—	—	—
Wisconsin	71.9	—	—	72.1	—	—
Wyoming	25.6	36.4	10.8	25.6	41.6	16.0 *
Average gain	†	†	7.0	†	†	9.9

— Not available.
† Not applicable.
* State and NAEP gains are significantly different from each other at $p<.05$.
NOTE: Primary standard is the state's standard for *proficient* performance.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.
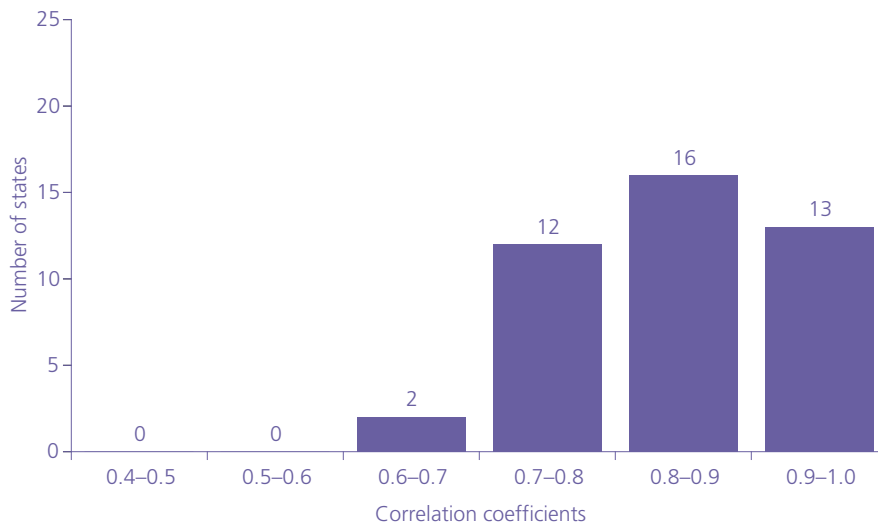
Table 9.   Mathematics achievement gains in percentage meeting the primary standard in grade 8, by state: 2000 and 2003

State/ jurisdiction	State			NAEP		
	2000	2003	Gain	2000	2003	Gain
Alabama	55.4	50.7	-4.7	41.1	39.0	-2.1
Alaska	—	65.2	—	—	65.2	—
Arizona	18.2	20.6	2.4	18.2	18.0	-0.2
Arkansas	13.3	21.9	8.6	13.3	18.1	4.8*
California	23.0	—	—	23.0	—	—
Colorado	—	68.2	—	—	68.2	—
Connecticut	57.2	56.0	-1.2	57.2	60.1	2.9*
Delaware	—	48.2	—	—	48.3	—
District of Columbia	8.6	—	—	8.5	—	—
Florida	—	54.1	—	—	54.2	—
Georgia	54.8	66.4	11.6	54.9	60.3	5.4*
Hawaii	60.5	54.1	-6.4	60.5	65.4	4.9*
Idaho	—	52.5	—	—	52.5	—
Illinois	45.8	53.6	7.8	45.8	48.7	2.9*
Indiana	—	—	—	—	—	—
Iowa	—	71.7	—	—	71.7	—
Kansas	55.3	59.3	4.0	55.3	55.6	0.3
Kentucky	26.0	31.9	5.9	26.0	29.9	3.9
Louisiana	7.3	8.7	1.4	7.3.0	10.5	3.2
Maine	20.6	17.1	-3.5	20.6	18.8	-1.8
Maryland	51.1	—	—	51.1	—	—
Massachusetts	33.7	38.1	4.4	33.7	41.1	7.4*
Michigan	—	—	—	—	—	—
Minnesota	—	—	—	—	—	—
Mississippi	—	46.0	—	—	46.1	—
Missouri	12.9	13.4	0.5	12.9	17.9	5.0*
Montana	—	70.3	—	—	70.4	—
Nebraska	59.7	—	—	60.9	—	—
Nevada	—	—	—	—	—	—
New Hampshire	—	—	—	—	—	—
New Jersey	—	56.2	—	—	56.1	—
New Mexico	—	39.4	—	—	39.3	—
New York	41.4	54.0	12.6	41.4	47.8	6.4*
North Carolina	80.7	82.2	1.5	80.8	84.1	3.3
North Dakota	—	43.6	—	—	43.6	—
Ohio	—	—	—	—	—	—
Oklahoma	70.5	71.3	0.8	70.5	73.5	3.0
Oregon	54.7	57.6	2.9	54.5	54.6	0.1
Pennsylvania	—	51.5	—	—	51.6	—
Rhode Island	26.5	35.3	8.8	26.6	28.7	2.1*
South Carolina	19.3	20.2	0.9	19.3	28.4	9.1*
South Dakota	—	57.5	—	—	57.5	—
Tennessee	—	57.7	—	—	42.6	—
Texas	89.8	—	—	89.9	—	—
Utah	57.9	56.8	-1.1	52.6	53.9	1.3
Vermont	45.6	51.7	6.1	45.7	52.4	6.7
Virginia	—	—	—	—	—	—
Washington	—	36.4	—	—	36.4	—
West Virginia	—	—	—	—	—	—
Wisconsin	42.9	—	—	43.0	—	—
Wyoming	31.5	35.1	3.6	31.6	41.3	9.7*
Average gain	†	†	3.0	†	†	3.6

— Not available.

† Not applicable.

* State and NAEP gains are significantly different from each other at $p<.05$.

NOTE: Primary standard is the state's standard for *proficient* performance.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

National Assessment of Educational Progress

SUMMARY

Comparisons are made between NAEP and state assessment mathematics achievement trends from 2000 to 2003. Achievement trends are measured by both NAEP and state assessments as gains in school-level percentages meeting the state's primary standard. Comparisons are based on the NAEP sample schools for which we also have state assessment scores. Trend data are available for 37 states. However, in eight of the states for which scores are available, the assessment and/or performance standards were changed during the period between 2000 and 2003; therefore, these states are not included in the trend analysis. As a result, comparisons of mathematics achievement trends from 2000 to 2003 are possible in 24 states for grade 4 and 22 states for grade 8.

In aggregate, in grade 4 but not grade 8, mathematics achievement gains from 2000 to 2003 reported by NAEP are significantly larger than those measured by state assessments. Gains measured by state assessments vary substantially between states; however, variability of gains measured by NAEP between states is only about two-thirds as large as the variation in state assessment results.

NAEP and state assessment gains in grade 4 mathematics are correlated positively. This indicates that, in grade 4 mathematics, states in which state assessments found the largest gains in percentages of students meeting the primary state standard tend to be the states in which NAEP results indicated the largest gains. However, grade 4 mathematics is the only grade and subject that showed a noticeable relationship between the NAEP gains and state assessment gains.

When comparisons between NAEP and state assessment 2000-2003 mathematics achievement trends are made for each state, significant differences are found in 14 out of the 24 states in grade 4 and 11 out of the 22 states in grade 8.

5 Comparing Achievement Gaps

A primary objective of federal involvement in education is to ensure equal opportunity for all students, including minority groups and those living in poverty (USDE 2002). NAEP has shown that although there have been gains since 1970, the average mathematics achievement of certain minority groups lag behind that of other students in both the elementary and secondary grades.[30] Numerous programs nationwide are aimed at reducing the mathematics achievement gap between Black and Hispanic students and White students, as well as between students in high-poverty and low-poverty schools; state assessments are monitoring achievement to determine whether, in their state, the gap is closing.

In compliance with *No Child Left Behind*, state education agencies now report school mathematics achievement results separately for minorities, for students eligible for free or reduced price lunch, for students with disabilities, and for English language learners (USDE 2002). These reports can be used to assess how successfully schools are narrowing the achievement gaps and to identify places needing assistance in narrowing their gaps.

Fair and unbiased measurement of the achievement of students from different cultural backgrounds is particularly difficult, and test developers try hard to remove test items that might unfairly challenge some groups more than others. In spite of these efforts, some state assessments may be more attuned to measuring achievement gaps and their narrowing than others. Comparison of NAEP measurement of mathematics achievement gaps to state assessment results can shed light on such differences.

The main objective of this part of the report is to compare the measurement of mathematics achievement gaps by state assessments and NAEP. Specifically, we compare three types of gaps:

- the Black-White achievement gap
- the Hispanic-White achievement gap
- the poverty gap: achievement of students qualifying for free or reduced-price lunch (i.e., disadvantaged students) versus those who do not qualify.[31]

---

30. Campbell, Hombo, and Mazzeo (2000).
31. We refer to students eligible for free/reduced price lunch as (economically) disadvantaged students.

The focus of these comparisons is not on differences in gaps between states but on differences between NAEP and state measurement of the gap in the same schools in the same state.

POPULATION PROFILES OF ACHIEVEMENT GAPS

Achievement gaps for whole subpopulations, such as Black students, Hispanic students, or economically disadvantaged students, are complex. What causes one segment of a disadvantaged population to achieve at a lower level may be quite different from the barriers faced by another segment of the same subpopulation. It is easy to forget that in the context of a population achievement gap, there are still many students in the disadvantaged group who achieve at a higher level than typical for non-disadvantaged groups. Expressing a mathematics achievement gap as a single number (the difference in the percentages of children in two groups who meet a standard) hides a great deal of information about the nature of gaps.

Moreover, as Paul Holland (2002) has shown, it is also likely to mislead readers because of the differential placement of the standard relative to the distribution of achievement in the two populations. For example, in figure 19 the poverty gap in mathematics achievement, which is about 10 points on the NAEP scale at the median, is larger in the lower part of the achievement distribution than in the higher part of the distribution. As a result, the gap is 13 percent in achieving the basic level (the distance between the points at which the graphs cross the basic criterion of 214 on the NAEP scale) and 13 percent in achieving the proficient level, but only 2 percent in achieving the advanced level. The graph, or population profile, conveys significantly more information about the poverty gap in mathematics than does a simple comparison of the percentages achieving the standards.

Holland points out that this effect is more striking when examining reduction in gaps. If we hypothetically suppose that at some future date all students would gain 20 points on the NAEP mathematics achievement scale, the population profiles would appear as in figure 20. In this case, the gaps in percent advanced, 8 percent, would be larger than in 2003, while the gap in percent achieving the basic level, 5 percent, would be smaller. Even though the achievement gaps remain constant, they appear to increase or decrease, depending on the position of the standards *vis-à-vis* the distribution of student achievement.

Even though the gap in the percentage of students achieving the basic level might be reduced from 13 percent to 5 percent, the gap in mathematics skills at that level would be just as large as before, larger than among higher-achieving segments of the disadvantaged and non-disadvantaged populations. Even though the gap in the percentage of students meeting the advanced standard might increase from 2 percent to 8 percent, it would be inappropriate to conclude that educators were ignoring the gap among the highest achievers in the two subpopulations.

Figure 19. Population profile of the NAEP poverty gap in grade 4 mathematics achievement: 2003
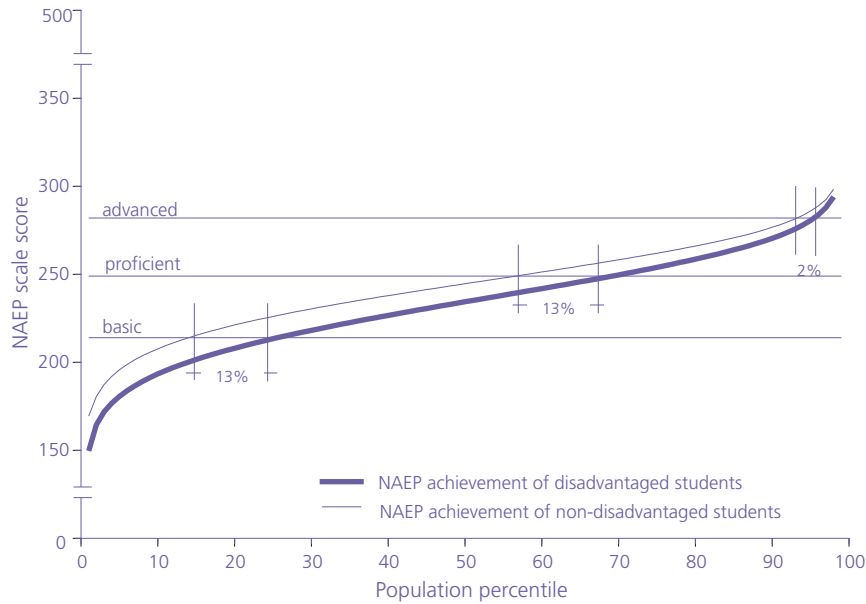
NOTE: Students eligible for free/reduced price lunch are referred to as (economically) disadvantaged.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates.

Figure 20. NAEP poverty gap from a hypothetical uniform 20-point increase on the grade 4 mathematics achievement scale
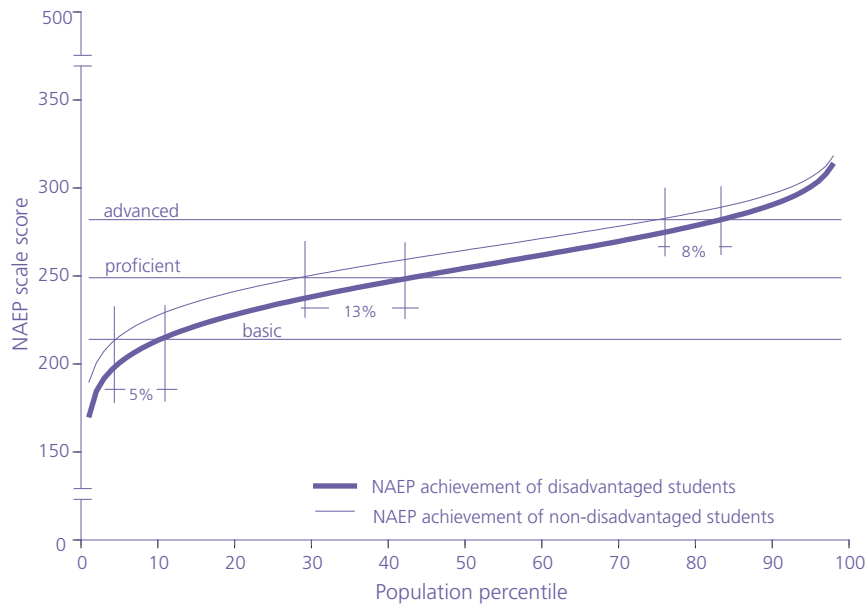
NOTE: Students eligible for free/reduced price lunch are referred to as (economically) disadvantaged.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates.

If gap measurement must be carried out in terms of percentages meeting standards, it is essential not to be misled by comparisons between gaps measured at different points on the achievement continuum. This effect makes it clear that comparison of NAEP and state assessment measurement of gaps in terms of percentages meeting standards must refer to the same standard for both assessments. Therefore, because individual scale values are not available for state assessments, we must measure gaps at the level of each state's primary standard.

It is important to note, however, that measuring the gap at each state's particular standard renders comparisons of gaps between states uninterpretable, because the standards are different in each state. And even though NAEP applies the same set of standards in all states to produce the biennial Nation's Report Card, comparisons of achievement gaps in different states must be interpreted in the context of variations in the position of the NAEP standards with respect to states' achievement distributions. Comparisons of gaps between states can be found in the Nation's Report Card (http://nces.ed.gov/nationsreportcard).

There are three major limitations in the data that further affect the interpretation of comparisons of gaps as measured by NAEP versus state assessments. The first limitation is that the state assessment data are only available at the school and grade level for the various population groups, not for individuals. The second is that percentages of subgroups meeting standards are suppressed in many schools, due to small samples. The third is that separate scores for subpopulations are not available in the NLSLSASD before 2002, precluding the possibility of comparing NAEP and state assessment of reduction in gaps between 2000 and 2003.

The state assessment data's limitation to school-level aggregate percentages of subgroups achieving standards means that each student is represented as the average of that student's population group in a school. As a result, variability in performance within each group in a school is not captured. The variability across schools using group averages is lower than the variability across schools using individual student scores. Unfortunately, to compare NAEP and state assessment gaps with each other, we must also limit the NAEP data to school averages for subgroups.

In addition, school-level scores are subject to *suppression* when the number of students tested in a subgroup is so small that the state prohibits the release of scores. Suppression rules vary between states; typically, scores are suppressed when fewer than 5 or 10 students are included in the average score. To avoid unreliable NAEP results in gap comparisons between NAEP and state assessment reports, we have omitted from analysis any schools with fewer than three tested subgroup members.[32]

---

32. Including percentages based on one or two students overestimates the frequency of observing extreme percentages: with one student, the percentage is either 0 or 100. Because small schools in the NAEP sample may be weighted to represent large numbers of small schools, this distorts some population profiles by overestimating the percentages of students in the extreme categories of 0 percent achieving the standard and 100 percent achieving the standard. Suppressing the cases based on one or two students more closely matches the state assessment practices of suppressing scores based on small sample sizes.

MATHEMATICS ACHIEVEMENT GAPS

The State Profile section of this report (appendix D) displays three types of achievement gaps similar to Figure 19, for all states with available data: the Black-White achievement gap, the Hispanic-White achievement gap, and the poverty achievement gap. These are introduced in the following pages through the presentation of population profiles of gaps showing the aggregation of percentages of students meeting states' primary standards. The graphs are not intended to reflect national average achievement gaps because they represent only some states and some schools, but they are informative. Although the graphs portray the aggregate achievement gap as measured against different standards across states, the general size of the gaps, in terms of standards in place in the nation in 2003, is apparent.

An aggregate population profile of the poverty gap in grade 8 mathematics achievement for the states included in this report is shown in the following series of four graphs, which compare mathematics achievement of disadvantaged students with other students. Figures 21 and 22 display the differential achievement as measured by NAEP and as measured by state assessments, respectively. [33]

In figure 21, the vertical (y) axis measures mathematics achievement. Due to limitations on the data available on state assessment results, mathematics achievement cannot be graphed for each individual student. Instead, it is measured by the percent of students in a school meeting the state's primary mathematics achievement standard. That is, for each student in a subgroup (such as disadvantaged students), the mathematics achievement measure is the percent of students in his or her subgroup in his or her school meeting the standard. Thus, within a particular school, all members of the subgroup have the same mathematics achievement measure, which is the average of all their individual achievement scores. The horizontal (x) axis represents all the students in the subgroup in the state or nation, arrayed from those with the lowest mathematics achievement measure on the left to the highest mathematics achievement measure on the right. The units on the horizontal axis are percentiles, from 0 to 100; that is, percentages of the student's subgroup with equal or lower mathematics achievement measures. Figure 21 serves a dual purpose. First, it arrays the disadvantaged student population (shown by the darker line) by percentile, from those in the schools where the fewest disadvantaged students meet the standard to those in schools where the most disadvantaged

---

33. A school at the 50th percentile on a population profile has an average performance that is higher than the average performance in schools serving half the students in the state and lower than the average performance in the schools serving the other half of the population (except for those in the school itself, of course). However, there may be different numbers of schools serving the upper and lower halves of the population. For example, for minority population profiles, there may be 500 (small) schools serving students in the upper half of the (minority) student population and 100 (larger) schools serving students in the lower half of that population, so the population percentile is not literally a percentile of schools. It is a percentile of the student population served by the schools.

students meet the standard. Second, it also arrays the non-disadvantaged student population (shown by the lighter line) by percentile. Thus, two population profiles can be superimposed on the same graph to display the achievement gap between them.

The population profiles in figure 21 can be read as follows. Focus first on comparing the highest achievers among disadvantaged students versus the highest achievers among non-disadvantaged students. Consider the 90th percentile of each population as an example. The vertical line at the 90th percentile crosses the *disadvantaged* line at 53 percent meeting the standard. That means that 90 percent of the population of disadvantaged students are attending schools where fewer than 53 percent of the disadvantaged students meet the standard and the other 10 percent are in schools where more than 53 percent of the disadvantaged students meet the standard. In other words, students at the *90th percentile of the disadvantaged population are attending schools where 53 percent of the disadvantaged students meet the standard*.

Figure 21. School percentage of economically disadvantaged and non-disadvantaged students meeting states' primary grade 8 mathematics standards as measured by NAEP, by percentile of students in each subgroup: 2003

NOTE: Primary standard is the state's standard for *proficient* performance. Students eligible for free/reduced price lunch are referred to as (economically) disadvantaged. Percentile in group refers to the percentage of the disadvantaged (or non-disadvantaged) student population who are in schools with lower (same-group) percentages meeting the states' primary mathematics standards.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

By comparison, the 90th percentile crosses the *non-disadvantaged* line in figure 21 at 79 percent meeting the standard. That means that 90 percent of *non-disadvantaged students* are attending schools where fewer than 79 percent of *non-disadvantaged students* meet the standard and the other 10 percent are in schools where more than 79 percent of *non-disadvantaged students* meet the standard. We can say that the *students at the 90th percentile of the non-disadvantaged population are attending schools where 79 percent of the non-disadvantaged students meet the standard*. Thus, comparing 90th percentile for the disadvantaged student population versus the 90th percentile for the non-disadvantaged student population, there is a gap of 26 points (26=79-53) in percentages between the groups in their school meeting the standard.

Similarly, at the tenth percentile, disadvantaged students are in schools where about 15 percent of disadvantaged students meet the standards. By comparison, non-disadvantaged students at the tenth percentile for their group are in schools where about 39 percent of non-disadvantaged students meet the standards, a gap of 24 percentage points among these relatively low-achieving students in the two groups.

Figure 22. School percentage of economically disadvantaged and non-disadvantaged students meeting states' primary grade 8 mathematics standards as measured by state assessments, by percentile of students in each subgroup: 2003

NOTE: Primary standard is the state's standard for *proficient* performance. Students eligible for free/reduced price lunch are referred to as (economically) disadvantaged. Percentile in group refers to the percentage of the disadvantaged (or non-disadvantaged) student population who are in schools with lower (same-group) percentages meeting the states' primary mathematics standards.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

The graphs in figures 21 and 22 are aggregated across the states and schools for which subgroup percentages achieving mathematics standards are available.[34] Because the primary standards vary from state to state, it is essential in comparing NAEP and state assessment results that the NAEP results are measured relative to each state's standard in that state. Corresponding population profiles of gaps for individual states are included in the State Profile section of this report (appendix D)—these aggregate profiles may provide context for interpreting individual state achievement gaps.

Figures 21 and 22 display similar pictures—throughout the middle range of the student populations, there is an apparently fairly uniform gap of between 20 and 30 percentage points, and this gap is noticeably smaller in the extreme high and low percentiles. Nevertheless, it should be clear that some disadvantaged students are in schools where the percentages of disadvantaged students achieving the standard are greater than the percentages of non-disadvantaged students achieving the standard in other schools. For example, figure 21 showed that the 90th-percentile disadvantaged students are in schools where 53 percent of disadvantaged students meet the standards, which is a greater percentage than among the lowest quarter of the non-disadvantaged student population.[35]

All of the graphs are based on the NAEP schools, weighted to represent the population of eighth graders in each state. Because the use of the aggregate school-level percentages, instead of individual student achievement scores, may have an effect on the position and shape of the population profile graphs, school-level percentages are presented in both NAEP and state assessment graphs. Appendix A presents a description of the method for constructing population profiles based on school-level aggregate achievement measures.

It is difficult to compare the NAEP and state assessment gaps when presented in different figures. Figure 23 combines the NAEP and state assessment profiles for disadvantaged and non-disadvantaged students. The similarity of the NAEP and state assessment results in this figure is notable. Although some discrepancies are worth noting, the overall picture suggests that as a summary across 34 states, NAEP and state assessments are measuring the poverty gap similarly.

Compared to the average state assessment results, the NAEP population profiles appear to exhibit greater variation between the top and bottom of the distributions; that is, there is greater variation in the achievement of mathematics standards

---

34. States with fewer than 10 percent disadvantaged students or fewer than 10 NAEP schools with non-suppressed percentages for disadvantaged students are excluded due to unstable estimates.

35. Readers should not be confused by the use of *percent* and *percentile* for the two axes in the population profile graphs. These are two completely different measures, which happen to have similar names. For percentiles, there must be a person at the lowest percentile and another at the highest percentile, by definition, because the percentiles just rank the people from lowest (zero, or one, in some definitions) to highest (100). The percent achieving a standard, on the other hand, can be zero for everybody (a very high standard indeed!) or 100 for everybody, or anywhere in between. The only built-in constraint is that the graphs must rise from left to right - higher achieving segments of the (sub)population are ranked in higher percentiles of the (sub)population by definition.

measured by NAEP, within both the disadvantaged and non-disadvantaged populations, than in achievement measured by state assessments (figure 23). Whether this is a real phenomenon or an artifact of the differences in design between NAEP and state assessments is not clear at this time and requires further study. One possibility is that because NAEP percentages meeting standards are generally based on fewer students in each school than state assessment percentages, variation of NAEP school means may naturally be larger than variation of state assessment school means.

In figure 23, information about the size of the difference between NAEP and state assessment measurement of the poverty gap is difficult to separate visually from information about the overall size of the gap. In particular, in the highest achievement percentiles, NAEP reports higher achievement by both groups than the aggregate state assessments do. To focus on the differences between NAEP and state assessments, we eliminate the distracting information by graphing only the *difference* between the profiles (the achievement of the disadvantaged group, minus the achievement of the non-disadvantaged group). The result is the gap profile in figure 24, in which a *zero* gap is the goal, and current gaps fall below that goal.

Figure 23. School percentage of economically disadvantaged and non-disadvantaged students meeting states' primary grade 8 mathematics standards as measured by NAEP and state assessments, by percentile of students in each subgroup: 2003

NOTE: Students eligible for free/reduced price lunch are referred to as (economically) disadvantaged. Percentile in group refers to the percentage of the disadvantaged (or non-disadvantaged) student population who are in schools with lower (same-group) percentages meeting the states' primary mathematics standards.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 24. Profile of the poverty gap in school percentage of students meeting states' grade 8 mathematics achievement standards, by percentile of students in each subgroup: 2003

NOTE: Students eligible for free/reduced price lunch are referred to as (economically) disadvantaged. Percentile in group refers to the percentage of the disadvantaged (or non-disadvantaged) student population who are in schools with lower (same-group) percentages meeting the states' primary mathematics standards.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

In figure 24 it becomes clearer how the NAEP and state assessment measurement of poverty gaps compare to each other. Although the gap profiles are similar, NAEP is measuring a larger gap in the lower three-quarters of the two populations. From the 25th to the 50th percentile, NAEP measures a gap of 28 to 30 percent in achieving state standards, while state assessments average a 24 to 26 percent gap. On the other hand, above the 75th percentile, the NAEP and average state assessment poverty gaps are virtually identical. Note that for individual state gap profiles (in appendix D) results of statistical significance tests are reported. Because the graphs presented in this section are not intended for inference about national patterns, no significance tests are reported here.

Similar population profiles can be constructed for other comparisons between subpopulations. We present these to provide a general and approximate national context in which to view the gaps displayed in appendix D for individual states. Figure 25 displays the grade 4 poverty gap corresponding to the grade 8 poverty gap shown in figure 24. The poverty gap profile for grade 4 is similar to that at grade 8, although the average discrepancy between NAEP and state assessment gaps is slightly larger, averaging about 5 percentage points between the 25th and 50th percentiles, compared to a 4 percentage point discrepancy for grade 8.

Figure 25. Profile of the poverty gap in school percentage of students meeting
states' grade 4 mathematics achievement standards, by percentile of
students in each subgroup: 2003

NOTE: Primary standard is the state's standard for *proficient* performance. Students eligible for free/reduced
price lunch are referred to as (economically) disadvantaged. Percentile in group refers to the percentage of the
disadvantaged (or non-disadvantaged) student population who are in schools with lower (same-group) per-
centages meeting the states' primary mathematics standards.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statis-
tics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population esti-
mates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

It should be noted that this comparison, like all comparisons between NAEP and
state assessment results in this report, is based on NAEP and state assessment results
in the same set of schools. For example, if the state reported a percentage meeting
their standard for disadvantaged students at a school, but the NAEP student sample
in that school included no disadvantaged students, that school would not be included
in the population profile of disadvantaged students. (Of course, that school might be
included in the non-disadvantaged student profiles: the gaps being reported here
combine between-school gaps as well as within-school gaps.)

Figures 26 and 27 provide aggregate population gap profiles for the Hispanic-White
gap and the Black-White gap in grade 8 mathematics achievement.[36] The Hispanic-
White gap is about 30 percent across much of the distribution as measured by NAEP,
slightly less when measured by state assessments, while the Black-White gap pattern
is about 36 to 37 percentage points as measured by NAEP, about 6 to 7 percent
greater than the state assessment measurement of the gap.

36. The aggregate Hispanic-White grade 8 mathematics achievement gap is based on results in 37
states, although sufficient sample sizes are available for comparison of results for individual states in
only 14 states. The aggregate Black-White gap is based on results in 39 states, although sufficient
sample sizes are available for comparison of results for individual states in only 20 states. The
aggregate poverty gap is based on results in 34 states, although sufficient sample sizes are available
for comparison of results for individual states in only 28 states.

Figure 26. Profile of the Hispanic-White gap in school percentage of students meeting states' grade 8 mathematics achievement standards, by percentile of students in each subgroup: 2003

NOTE: Primary standard is the state's standard for *proficient* performance. Percentile in group refers to the percentage of the Hispanic or White student population who are in schools with lower (same-group) percentages meeting the states' primary mathematics standards.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 27. Profile of the Black-White gap in school percentage of students meeting states' grade 8 mathematics achievement standards, by percentile of students in each subgroup: 2003

NOTE: Primary standard is the state's standard for *proficient* performance. Percentile in group refers to the percentage of the Black or White student population who are in schools with lower (same-group) percentages meeting the states' primary mathematics standards.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Corresponding aggregate grade 4 Hispanic-White and Black-White mathematics achievement gap profiles are shown in figures 28 and 29. While the grade 4 gap comparison results are similar to the grade 8 comparisons, in grade 4, state assessments tend to find a slight narrowing of the gap between the upper thirds of the populations, compared to the lower two-thirds of the populations, a pattern that is not seen in the grade 8 comparisons.

Although the aggregate gap profile can identify small reliable differences in gaps as measured by NAEP and state assessments, the samples in individual states are not sufficiently large to detect small differences. The differences between the NAEP and state assessments in mean state gaps in percentage meeting the primary grade 4 mathematics standards are shown in table 10 for the states in which sufficient subgroup data are available. Although these statistics are based on the overall average, readers can examine Student's *t* test results for various parts of the distributions (halves and quartiles) for individual states in appendix D.

At grade 4, in about one-half of the individual state comparisons (34 of 70), the overall gap as measured by NAEP in 2003 was statistically significantly larger than the gap measured by the state assessment, and in all but two of the other 36 comparisons, the NAEP gap was numerically greater (although not significantly so). NAEP found significantly larger mean gaps than the state assessment did in 14 of 25 Black-White comparisons, in Alabama, Arkansas, Connecticut, Florida, Georgia, Indiana, Kansas, Louisiana, Massachusetts, Missouri, New Jersey, Ohio, South Carolina, and Virginia.

Of 14 Hispanic-White comparisons, NAEP gaps were larger in eight: Arizona, California, Idaho, Illinois, Nevada, New Jersey, New Mexico, and Rhode Island; and of 31 poverty comparisons, NAEP gaps were larger in 12: Alabama, California, Connecticut, District of Columbia, Hawaii, Indiana, Nevada, New Hampshire, Ohio, Vermont, Wisconsin, and Wyoming. Note that Black-White gap results for Kansas and Missouri and poverty gap results for New Hampshire and Vermont might be affected by the fact that the schools available for comparison represented low percentages of the subgroup populations.

At grade 8, as displayed in table 11, NAEP and state assessment measurements of gaps were somewhat more similar to each other. However, in 17 of 62 comparisons, NAEP found a significantly larger gap than the state assessment did, while in only two cases did the state assessment find a significantly larger gap. NAEP found significantly larger Black-White gaps in seven of 20 state comparisons, in Alabama, Florida, Georgia, Mississippi, New York, Texas, and Virginia; significantly larger Hispanic-White gaps in six of 14 states, Arizona, Florida, Idaho, Nevada, New Mexico, and Rhode Island; and significantly larger poverty gaps in four of 28 states, Alabama, Georgia, Hawaii, and Illinois. By contrast, the state's assessments found a larger poverty gap than NAEP did in South Carolina and Wyoming.

Figure 28. Profile of the Hispanic-White gap in school percentage of students meeting states' grade 4 mathematics achievement standards, by percentile of students in each subgroup: 2003

NOTE: Primary standard is the state's standard for *proficient* performance. Percentile in group refers to the percentage of the Hispanic or White student population who are in schools with lower (same-group) percentages meeting the states' primary mathematics standards.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 29. Profile of the Black-White gap in school percentage of students meeting states' grade 4 mathematics achievement standards, by percentile of students in each subgroup: 2003

NOTE: Primary standard is the state's standard for *proficient* performance. Percentile in group refers to the percentage of the Black or White student population who are in schools with lower (same-group) percentages meeting the states' primary mathematics standards.
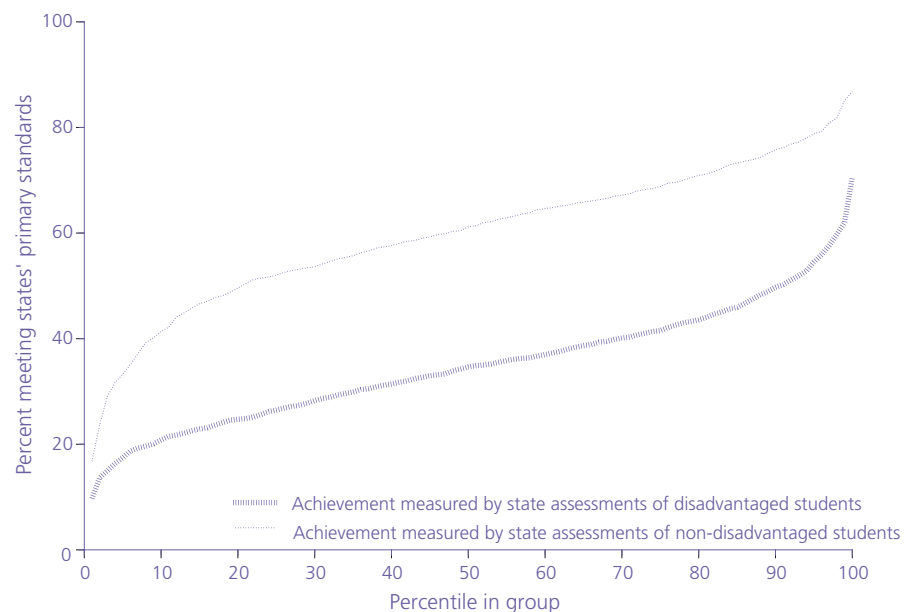
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table 10. Differences between NAEP and state assessments of grade 4
mathematics achievement race and poverty gaps, by state: 2003

State/ jurisdiction	Black-White	Hispanic-White	Poverty
Alabama	-7.5 *	—	-4.2 *
Alaska	—	—	—
Arizona	—	-12.1 *	—
Arkansas	-8.5 *	—	0.5
California	—	-11.1 *	-6.7 *
Colorado	—	—	—
Connecticut	-9.5 *	-5.4	-5.5 *
Delaware	-1.4	—	0.5
District of Columbia	—	—	-15.8 *
Florida	-8.2 *	-3.8	-3.4
Georgia	-8.4 *	—	-3.9
Hawaii	—	—	-5.7 *
Idaho	—	-10.4 *	—
Illinois	-2.5	-7.1 *	-3.7
Indiana	-22.7 *	—	-11.5 *
Iowa	—	—	—
Kansas	-11.2 *	—	-7.7
Kentucky	-1.3	—	-2.3
Louisiana	-4.7 *	—	-3.2
Maine	—	—	—
Maryland	—	—	—
Massachusetts	-5.4 *	-6.7	—
Michigan	—	—	—
Minnesota	—	—	-2.9
Mississippi	-1.1	—	-3.7
Missouri	-13.3 *	—	-3.7
Montana	—	—	—
Nebraska	—	—	—
Nevada	—	-7.2 *	-7.1 *
New Hampshire	—	—	-10.8 *
New Jersey	-8.5 *	-14.7 *	-5.6
New Mexico	—	-8.5 *	-2.9
New York	-4.1	-5.0	-1.3
North Carolina	-4.8	—	-2.1
North Dakota	—	—	—
Ohio	-8.3 *	—	-8.2 *
Oklahoma	-5.5	—	—
Oregon	—	—	—
Pennsylvania	-4.8	—	-3.3
Rhode Island	—	-13.9 *	—
South Carolina	-5.5 *	—	-2.9
South Dakota	—	—	-3.7
Tennessee	-3.1	—	-0.3
Texas	-2.9	-1.1	—
Utah	—	—	—
Vermont	—	—	-14.3 *
Virginia	-10.3 *	—	—
Washington	—	-0.2	—
West Virginia	—	—	—
Wisconsin	-8.1	—	-9.5 *
Wyoming	—	—	-4.8 *

— Not available.

* NAEP-state difference is statistically significantly at *p*<.05.

NOTE: A positive entry indicates that the state assessment reports the gap as larger than NAEP does; a negative entry indicates that the state assessment reports the gap as smaller.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table 11. Differences between NAEP and state assessments of grade 8
mathematics achievement race and poverty gaps, by state: 2003

State/ jurisdiction	Black-White	Hispanic-White	Poverty
Alabama	-8.1 *	—	-3.6 *
Alaska	—	—	—
Arizona	—	-5.4 *	—
Arkansas	0.9	—	4.1
California	—	-4.6	-2.1
Colorado	—	—	—
Connecticut	0.2	1.6	0.6
Delaware	-2.3	—	1.3
District of Columbia	—	—	-1.3
Florida	-6.8 *	-6.3 *	-1.7
Georgia	-13.1 *	—	-7.5 *
Hawaii	—	—	-4.6 *
Idaho	—	-6.4 *	—
Illinois	-1.3	-4.9	-5.6 *
Indiana	-0.5	—	4.2
Iowa	—	—	—
Kansas	—	—	1.5
Kentucky	—	—	-0.2
Louisiana	-1.1	—	-0.7
Maine	—	—	—
Maryland	—	—	—
Massachusetts	—	-4.3	—
Michigan	—	—	—
Minnesota	—	—	—
Mississippi	-6.9 *	—	-4.5
Missouri	-0.8	—	-0.6
Montana	—	—	—
Nebraska	—	—	—
Nevada	—	-5.1 *	-0.0
New Hampshire	—	—	—
New Jersey	-1.7	-2.0	-2.1
New Mexico	—	-7.2 *	-0.7
New York	-9.8 *	-5.9	-1.6
North Carolina	-4.5	—	-1.1
North Dakota	—	—	—
Ohio	—	—	—
Oklahoma	-0.7	—	—
Oregon	—	-5.2	—
Pennsylvania	-4.6	—	-3.3
Rhode Island	—	-3.3 *	—
South Carolina	3.5	—	4.7 *
South Dakota	—	—	-0.3
Tennessee	-2.7	—	3.6
Texas	-7.3 *	-2.4	—
Utah	—	—	—
Vermont	—	—	-4.2
Virginia	-8.6 *	—	—
Washington	—	—	—
West Virginia	—	—	—
Wisconsin	—	—	-1.6
Wyoming	—	—	4.6 *

— Not available.
* NAEP-state difference is statistically significantly at $p<.05$.
NOTE: A positive entry indicates that the state assessment reports the gap as larger than NAEP does; a negative entry indicates that the state assessment reports the gap as smaller.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Not very much should be made of these significant results, until additional studies are performed. Examination of the individual state gap profiles in appendix D supports the conclusion that for the most part, measurement of gaps by NAEP and by state assessments were qualitatively similar. There are only two cases in which we were able to carry out comparisons and in which the NAEP gap is more than twice as large as the state assessment gap: these were the grade 4 Black-White gap in Indiana and poverty gap in the District of Columbia. Various factors, both substantive and methodological, may explain the tendency for NAEP to find slightly larger gaps where differences were found.[37] These must be factors that differentially affect the measurement of performance of students in different groups.

Among such possible factors, on the methodological side, there could be differences in student motivation, in methods of analyzing the test scores, or in prevalence of testing accommodations. Similarly, on the substantive side, it is possible that variation in scores on a state assessment, which focuses on what is taught in the schools, is somewhat less related to cultural differences that children bring to their schoolwork, compared to NAEP, because NAEP aims for an overall assessment of mathematics achievement, including both school- and culturally-related components of that performance.

SUMMARY

Comparisons are made between NAEP and state assessment measurement of mathematics achievement gaps in grades 4 and 8 in 2003. Comparisons are based on school-level percentages of Black, Hispanic, White, and economically disadvantaged and non-disadvantaged students achieving the state's primary mathematics achievement standard in the NAEP schools in each state. In most states, the comparison is based on state test scores for grades 4 and 8, but scores from adjacent grades are used for the comparisons in a few states. Comparisons of gaps are subject to data availability. Black-White and poverty comparisons for 2003 are possible in 31 states, and Hispanic-White comparisons are possible in 14.

In about half of the states, gap profiles based on NAEP show somewhat larger gaps than profiles based on state assessments. Of 132 state comparisons (composed of Black-White Hispanic-White, grade 4, and grade 8 comparisons across multiple states), NAEP found significantly larger gaps in 51 and state assessments found larger gaps in 2. These results contrast with results for reading, where NAEP found significantly larger gaps in 13 comparisons, while state assessments found larger gaps in 8 comparisons.

---

37. The determination of those factors is beyond the scope of the report.

6 Supporting Statistics

I n this section, we address two of the issues that were faced in preparing to compare NAEP and state assessment results: (1) the changing rates of exclusion and accommodation in NAEP; and (2) the effects of using the NAEP sample of schools for the comparisons.

STUDENTS WITH DISABILITIES AND ENGLISH LANGUAGE LEARNERS

Many factors affect comparisons between NAEP and state assessment measures of mathematics achievement trends and gaps. One of these factors is the manner in which the assessments treat the problem of measuring the mathematics achievement of students with disabilities (SD) and English language learners (ELL). Before the 1990s, small percentages of students were excluded from testing, including national NAEP as well as state assessments. In the 1990s, increasing emphasis was placed on providing equal access to educational opportunities for SD and ELL students, including large-scale testing (Lehr and Thurlow, 2003). Both NAEP and state assessment programs developed policies for accommodating the special testing needs of SD/ELL students to decrease the percentages of students excluded from assessment.

In the period since 1995, NAEP trends have been subject to variation due to changing exclusion rates in different states (McLaughlin 2000, 2001, 2003). Because that variation confounds comparisons between NAEP and state assessment results, the NAEP computations in this report have been based on full population estimates (FPE). The full population estimates incorporated questionnaire information about the differences between included and excluded SD/ELL students in each state to impute plausible values for the students excluded from the standard NAEP data files (McLaughlin 2000, 2001, 2003; Wise et al., 2004). Selected computations ignoring the subpopulation of students represented by the roughly 5 percent of students excluded from NAEP participation are presented in appendix C. Later in this section, we also compare (in tables 14, 15, 16, and 17) the average differences in the results obtained by using the full population estimates versus those results obtained when we used the standard NAEP estimates.

Research on the effects of exclusions and accommodations on assessment results has not yet identified their impact on gaps and trends. However, to facilitate exploration of possible explanations of discrepancies between NAEP and state assessment results in terms of exclusions and accommodations, table 12 displays NAEP estimates of percentages of the population identified, excluded, and accommodated in 2000 and 2003 for grades 4 and 8.

Table 12. Percentages of grades 4 and 8 English language learners and students with disabilities identified, excluded, or accommodated in NAEP mathematics assessments: 2000 and 2003

	Grade 4		Grade 8	
Students	**2000[1]**	**2003**	**2000[1]**	**2003**
Identified	**19.0**	**22.2**	**16.9**	**18.5**
Students with disabilities	10.7	11.7	11.2	12.1
English language learners	7.4	8.5	4.9	4.7
Both	0.9	2.0	0.9	1.6
Excluded	**4.2**	**3.9**	**4.4**	**3.8**
Students with disabilities	2.7	2.4	3.2	2.6
English language learners	1.1	0.9	0.8	0.7
Both	0.4	0.6	0.4	0.5
Accommodated	**6.7**	**8.3**	**4.1**	**6.8**
Students with disabilities	4.4	6.1	3.1	5.6
English language learners	2.1	1.5	0.7	0.7
Both	0.2	0.7	0.2	0.5

1. Alaska, Colorado, Delaware, Florida, New Hampshire, New Jersey, Pennsylvania, South Dakota, and Washington are not included in totals, and Iowa is not included for grade 8.

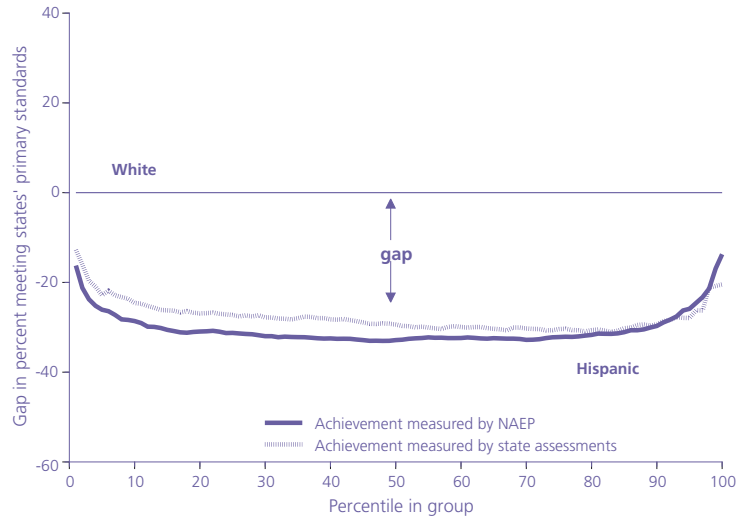NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments.

The top segment of table 12 displays the percentages of students identified as SD, ELL, and both, in recent NAEP mathematics assessments. The percentages shown for SD and ELL do not include students identified with both special needs, so each total is the sum of the three subgroups. These percentages include students subsequently excluded from participation, and they are weighted to represent percentages of the student population. The figures are aggregated over the states participating in NAEP at the state level in each case.[38] Individual state figures are displayed in the State Profiles section of this report (Appendix D).

The middle segment of table 12 displays the percentages of students who were excluded from participation in the NAEP test sessions. As before, these figures represent percentages of the student population. The bottom segment of the table displays the percentages of students who were provided with testing accommodations.

---

38. For 2000, Alaska, Colorado, Delaware, Florida, New Hampshire, New Jersey, Pennsylvania, South Dakota, and Washington are not included in totals, and Iowa is not included for grade 8.

Students identified as SD outnumber those identified as ELL by a factor of 3 to 2 in grade 4 and a factor of approximately 2 to 1 at grade 8. There was a 10 percent or more increase in the aggregate percentage of students identified as either SD or ELL between 2000 and 2003: from 19 to 22 percent at grade 4 and from 17 to 19 percent at grade 8. These percentages and their changes varied substantially between states, as shown in tables in appendix D.

While the figures in table 12 emphasize that the percentages of students who were excluded and accommodated were a small fraction of the students selected to participate in NAEP, they do not show the actual rates of exclusion of students with disabilities and English language learners. Table 13 displays these rates, along with the rates at which students with disabilities and English language learners who are included are provided with testing accommodations.

Table 13. Percentages of those identified as English language learner or as with disabilities, excluded, or accommodated in the NAEP mathematics assessments grades 4 and 8: 2000 and 2003

	Grade 4		Grade 8	
Students identified	**2000[1]**	**2003**	**2000[1]**	**2003**
Excluded	**22.3**	**17.6**	**26.0**	**20.3**
Students with disabilities	25.7	20.6	28.3	21.4
English language learners	14.8	10.6	16.8	14.6
Both	44.1	28.8	48.9	28.2
Accommodated	**45.5**	**45.4**	**32.8**	**46.2**
Students with disabilities	54.9	65.2	39.2	58.9
English language learners	33.8	20.4	18.4	18.4
Both	42.0	49.3	47.6	38.6

1. Alaska, Colorado, Delaware, Florida, New Hampshire, New Jersey, Pennsylvania, South Dakota, and Washington are not included in totals, and Iowa is not included for grade 8.

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments.

At grade 4 in 2000, 22 percent of the students identified as SD or ELL were excluded from NAEP, but this percentage was reduced to fewer than 18 percent in 2003. In 2000 and 2003, a smaller percentage of identified students were excluded from mathematics sessions in grade 4 than in grade 8.

NAEP has gradually increased its permission rules and procedures for the use of testing accommodations for SD and ELL, in an effort to reduce exclusions. By 2000, nearly one half of SD and ELL students participating in grade 4 NAEP sessions were provided accommodations, and this remained constant through 2003. However, in grade 8, the percentage increased from about one third to nearly one half between 2000 and 2003. There is little research to address the question of how that increase affects the measurement of trends.

NAEP Full Population Estimates and Standard Estimates

In this report, unlike previous NAEP reports, achievement estimates based on questionnaire and demographic information for this subpopulation are incorporated in the NAEP results. NAEP statistics presented are based on full population estimates, which include imputed performance for students with disabilities and English language learners who are excluded from participation in NAEP. As shown in table 12, these are roughly 4 percent of the students selected to participate in NAEP. Standard NAEP estimates do not represent this 4 percent of the student population, whose average mathematics achievement is presumably lower than the mathematics achievement of the 96 percent of students included in the standard NAEP estimates. Because the percentages of students excluded by NAEP vary from state to state, from year to year, and between population subgroups, estimates of trends and gaps can be substantially affected by exclusion rates. While we have not been able to adjust for varying exclusion rates in state assessment data in this report, we have, for the most part, eliminated the effects of varying exclusion rates in the NAEP data.

The method of imputation is based on information from a special questionnaire completed for all SDs and ELLs selected for NAEP, whether or not they are excluded. The method of imputation is described in appendix A. The basic assumption of the imputation method is that excluded SDs and ELLs with a particular profile of teacher ratings and demographics would achieve at the same level as the included SDs and ELLs with the same profile of ratings and demographics in the same state.

All comparisons between NAEP and state assessment results in this report were carried out a second time using the standard NAEP estimates. Four tables (tables 14-17) below summarize the comparisons of mathematics standards, correlations, trends, and gap computations we derived by using the full population estimates (FPE), versus the standard NAEP estimates (SNE). The summary figures in these tables (unweighted averages, standard deviations, and counts of statistically significant differences) are based on the individual state results presented in tables in the preceding sections, which are full population estimates, and standard NAEP estimates presented in appendix C.

Table 14 below shows the average differences in the NAEP equivalents of primary state mathematics standards in 2003. Although the FPE-based NAEP equivalents were about one point lower than SNE-based equivalents, due to inclusion of more low achieving students in the represented population, there was noticeable variation between states, due to variations in NAEP exclusion rates between states

Table 14. Difference between the NAEP score equivalents of primary mathematics achievement standards, obtained using full population estimates (FPE) and standard NAEP estimates (SNE), by grade: 2003

Level	Number of states	Mean difference of NAEP equivalent standards: FPE-SNE	Standard deviation of difference
Grade 4	46	-1.1	0.7
Grade 8	43	-1.1	0.8

NOTE: Primary standard is the state's standard for *proficient* performance. Negative mean differences in the NAEP equivalent standards indicate that the standards based on full population estimates are lower than the standards based on standard NAEP estimates.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-level State Assessment Score Database (NLSLSASD) 2004.

Table 15 shows that there were virtually no differences in the correlations (of the NAEP and state assessment percentages meeting grades 4 and 8 mathematics standards in 2003) between those correlations computed using the full population estimates (presented in table 4) and the standard NAEP estimates (in table C3).

Table 15. Difference between correlations of NAEP and state assessment school-level percentages meeting primary state mathematics standards, obtained using NAEP full population estimates (FPE) and standard NAEP estimates (SNE), by grade: 2003

Level	Number of states	Mean difference of NAEP equivalent standards: FPE-SNE	Standard deviation of difference
Grade 4	49	0.00	0.01
Grade 8	46	0.00	0.01

NOTE: Primary standard is the state's standard for *proficient* performance. Positive mean differences indicate that the correlations based on the full population estimates are greater than the correlations based on the standard NAEP estimates. For three states, the correlated achievement measure is the median percentile rank.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-level State Assessment Score Database (NLSLSASD) 2004.

Table 16 shows the average differences in the trends of 4th and 8th grade mathematics performance from 2000 to 2003 when those trends are computed (a) using the full population estimates (presented in table 8 and table 9) and (b) using the standard NAEP estimates (presented in tables C4 and C5). There is no mean difference in trends, and although the differences varied somewhat from state to state, no differences were sufficient to change the result of a test for statistical significance.

Table 16. Mean difference in mathematics performance gains between 2000 and 2003, based on NAEP full population estimates (FPE) versus standard NAEP estimates (SNE), by grade

| Level | Number of states | Mean difference in gain FPE-SNE | | Standard deviation of difference in gains: FPE-SNE | | Number of statistically significant differences between NAEP and state assessment gains | |
		State assessment	NAEP	State assessment	NAEP	FPE	SNE
Grade 4	24	0.0	0.0	0.2	0.7	14	14
Grade 8	22	0.0	0.0	0.2	0.6	11	11

NOTE: Positive mean differences in the NAEP equivalent standards indicate that the gains based on full population estimates are larger, though not always significantly, than the gains based on standard NAEP estimates.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-level State Assessment Score Database (NLSLSASD) 2004.

Finally, we compare the differences between full population estimates and standard NAEP estimates on gap comparisons. Table 17 shows the average differences in the achievement gaps (in 4th and 8th grade mathematics performance) between those gaps computed using the full population estimates (presented in table 10 and table 11) and the achievement gaps computed using the NAEP reported data (presented in table C6). The figures in tables 10 and 11 and C6 are differences between the gaps as measured by NAEP and the gaps as measured by state assessments. A positive entry in those tables indicated that the NAEP measure of the gap was smaller than the state assessment of the gap. For table 17, we subtract the NAEP-state assessment differences based on standard NAEP estimates from the NAEP-state assessment differences based on full population estimates.

Table 17. Mean difference in gap measures of mathematics performance obtained using NAEP full population estimates (FPE) versus standard NAEP estimates (SNE), by grade: 2003

| Level | Gap | Number of states | Mean difference in gaps: FPE-SNE | Standard deviation of difference in gaps: FPE-SNE | Number of statistically significant differences between NAEP and state assessment gaps | |
					FPE	SNE
	Black–White	25	0.8	0.8	14	17
Grade 4	Hispanic–White	14	-0.2	0.7	8	8
	Poverty	31	-0.2	0.6	12	12
	Black–White	20	0.2	0.5	7	9
Grade 8	Hispanic–White	14	0.0	0.5	6	5
	Poverty	28	-0.4	0.6	6	6

NOTE: Positive mean differences indicate that NAEP finds smaller gaps than state assessments to a greater extent when using the full population estimates than when using standard NAEP estimates.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-level State Assessment Score Database (NLSLSASD) 2004.

Overall, the gap comparison results for standard NAEP estimates are similar to the results for full population estimates. Both sets of estimates agreed that NAEP measures of gaps were significantly larger than state assessment measures of those gaps in 49 comparisons and state assessments of gaps were larger in two comparisons. In only eight cases did the two sets of estimates disagree; and these disagreements did not exhibit any bias for one or the other method to estimate NAEP gaps to be relatively larger or smaller than state assessment gaps (four in each direction).

In summary, for measurement of gains in mathematics achievement since 2000 and minority and poverty gaps in mathematics achievement, as well as for correlations of percentages meeting mathematics standards between NAEP and state assessments, the choice to use full population estimates or standard NAEP estimates has only minor effects on the outcomes of comparisons between NAEP and state assessment results. That is, changes in exclusion rates between the 2000 and 2003 NAEP mathematics assessments and differences in exclusion rates between subpopulations had only minor effects on these NAEP-state assessment comparisons. That does not imply, it should be pointed out, that the use of these two different methods would yield the same results in comparisons of NAEP mathematics achievement gains and gaps between states, comparisons not undertaken in this report.

USE OF SCHOOL-LEVEL DATA FOR COMPARISONS BETWEEN NAEP AND STATE ASSESSMENT RESULTS

One of the critical issues for NAEP-state assessment comparisons is whether the comparisons are based on the same populations. In order to ensure that differences that might be found between NAEP and state assessment results would not be attributable to different sets of schools, our comparisons were carried out on schools in the NAEP sample, and summary state figures were constructed from the results in those schools, using NAEP weights. One barrier to this approach was the challenge of finding the state assessment scores for the several thousand schools participating in each of the NAEP assessments. In this section, we present information on that matching process. In addition, as a validation of both the NAEP sample and the match between (a) the state assessment data on the databases we used and (b) the data used by the states for their reports, we compare our estimates of the percentages of students meeting state standards with the percentages reported on state websites.

State assessment results for NAEP schools
Our aim was to match state assessment scores to all of the public schools participating in NAEP. The percent of schools matched for the 2003 NAEP assessments are displayed in table 18. At grade 4, the median match rate across states was 99.1 percent. That is, of the approximately 100 schools per state per assessment, we found state assessment records for all, or all but one, in most states. The fact that the median weighted match rate was over 99 percent indicates that the schools we missed tended to be schools carrying less weight in computing state averages from the NAEP

Table 18. Weighted and unweighted percentages of NAEP schools matched to state assessment records in mathematics, by grade and state: 2003

State/jurisdiction	Grade 4		Grade 8	
	unweighted	weighted	unweighted	weighted
Alabama	99.1	97.7	99.0	99.2
Alaska	100.0	100.0	100.0	100.0
Arizona	93.4	91.7	96.6	96.1
Arkansas	100.0	100.0	100.0	100.0
California	99.2	99.0	99.5	99.9
Colorado	96.8	96.0	98.2	98.3
Connecticut	100.0	100.0	100.0	100.0
Delaware	92.0	91.7	97.3	98.4
District of Columbia	87.3	89.1	73.7	83.4
Florida	98.1	98.3	99.0	99.1
Georgia	96.2	96.9	96.6	95.4
Hawaii	100.0	100.0	83.6	98.5
Idaho	100.0	100.0	100.0	100.0
Illinois	99.4	99.5	100.0	100.0
Indiana	100.0	100.0	100.0	100.0
Iowa	97.8	98.6	98.3	98.5
Kansas	99.3	98.3	99.2	99.6
Kentucky	100.0	100.0	100.0	100.0
Louisiana	100.0	100.0	100.0	100.0
Maine	98.7	99.9	99.1	99.8
Maryland	100.0	100.0	100.0	100.0
Massachusetts	100.0	100.0	100.0	100.0
Michigan	99.3	99.6	100.0	100.0
Minnesota	99.1	99.9	100.0	100.0
Mississippi	99.1	99.0	100.0	100.0
Missouri	100.0	100.0	100.0	100.0
Montana	99.4	99.9	100.0	100.0
Nebraska	90.4	97.9	91.3	98.2
Nevada	98.2	96.3	97.0	96.4
New Hampshire	99.2	99.1	86.9	86.1
New Jersey	100.0	100.0	100.0	100.0
New Mexico	98.3	98.8	93.8	94.6
New York	98.0	98.3	97.3	98.2
North Carolina	98.7	99.6	98.5	98.4
North Dakota	97.6	99.6	100.0	100.0
Ohio	98.2	92.3	75.2	66.7
Oklahoma	100.0	100.0	99.2	99.5
Oregon	99.2	98.8	100.0	100.0
Pennsylvania	90.4	89.4	100.0	100.0
Rhode Island	99.1	99.6	98.1	98.4
South Carolina	97.2	98.2	99.0	98.7
South Dakota	89.8	98.4	89.1	98.8
Tennessee	100.0	100.0	99.1	96.5
Texas	99.0	97.8	98.6	95.9
Utah	98.3	96.9	100.0	100.0
Vermont	99.4	97.2	100.0	100.0
Virginia	95.7	93.8	96.3	94.0
Washington	99.1	99.7	100.0	100.0
West Virginia	100.0	100.0	100.0	100.0
Wisconsin	100.0	100.0	100.0	100.0
Wyoming	97.6	99.7	93.3	99.4

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

sample. The overall success of the matching process was equally good at grade 8, where the median match rate was 99.2 percent, with a median weighted match rate of 99.8 percent.

For grade 4, the only jurisdiction with a matching rate less than 90 percent were the District of Columbia (87 percent) and South Dakota (90 percent). In South Dakota, some of the unmatched schools are likely to be small schools for which all state assessment scores are suppressed. Schools having all missing data for assessment results in state assessment files had purposefully been excluded from the NLSLSASD, the database from which we extracted state assessment information for this report. These tended to be small schools, which are more prevalent in rural states such as South Dakota. The weighted match rate for South Dakota was 98.4 percent.

For grade 8, we were able to match more than 90 percent of the schools in all but five jurisdictions: District of Columbia (74 percent), Ohio (75 percent), Hawaii (84 percent), New Hampshire (87 percent), and South Dakota (89 percent). For Ohio, we do not include any grade 8 results in this report; and for Hawaii and South Dakota, the weighted match rates are very high. However, for the District of Columbia and New Hampshire, the lower match rate may offer one explanation for any discrepancies that are found between NAEP and state testing results for grade 8.

Failure to match a NAEP school to the state records is not the only source of omission of NAEP schools from the comparison database. As indicated in table 1, the percentages of schools used for analyses were somewhat lower in certain states. In many states, the percentages of the population represented in the analyses clustered around 90 percent; however the comparison samples in Arizona, Delaware, New Mexico, and Tennessee included schools that represented less than 85 percent of the NAEP sample at grade 4. At grade 8, more than 85 percent of the student population was represented in the analyses for all jurisdictions except the District of Columbia, New Mexico, North Dakota, Tennessee, and Washington.

Failure to match all NAEP schools is not likely to have a significant impact on the comparison analyses unless the missing schools are systematically different from other schools. In fact, due to suppression of state assessment scores for small reporting samples missing schools in these analyses, missing schools are more likely to be small schools. Interpretation of the findings should take this potential bias into account.

This is an even more critical issue with respect to the gap analyses, where small to moderate-sized schools with small percentages of minority students are more likely to have their minority average achievement scores suppressed. And to balance the gap analyses, schools with only one or two NAEP minority participants were excluded from the minority population used to construct the population achievement profile for that minority. The percentages of the minorities represented by the NAEP data that are included in gap analyses in each state are displayed in table 19.

Across the states for which gap profiles are included in this report, the median percentages of Hispanic students and disadvantaged students included in grade 4 analyses is 85 percent, and the median percentage of Black students is 87 percent. In

most states, more than two-thirds of the minority students are included, and in all states, more than half are included. The states with fewer than two-thirds of Black students included are Connecticut, Delaware, Kansas, Missouri, and Wisconsin. Connecticut and Idaho Hispanic gap analyses are based on fewer than two-thirds of the Hispanic students in each state; and poverty gap analyses in Delaware, Missouri, New York, Vermont, and New Hampshire are based on fewer than two-thirds of the disadvantaged students in these states, based on NAEP estimates.

At grade 8, the situation is better, because with larger schools, fewer minority data are suppressed in state assessment files. The median percentages included in gap analyses are 94 percent for Blacks, 92 percent for Hispanics, and 90 percent for disadvantaged students; and there are no states in which analyses are based on fewer than 70 percent of the minority students in NAEP files.

Table 19. Percentages of NAEP student subpopulations in grades 4 and 8 included in comparison analysis in mathematics, by state: 2003

State/ jurisdiction	Black students		Hispanic students		Disadvantaged students	
	Grade 4	Grade 8	Grade 4	Grade 8	Grade 4	Grade 8
Alabama	92.8	90.1	—	—	93.2	89.7
Alaska	—	—	—	—	—	—
Arizona	—	—	77.5	90.7	—	—
Arkansas	95.8	84.4	—	—	98.1	92.6
California	—	—	94.3	98.0	94.9	97.6
Colorado	—	—	—	—	—	—
Connecticut	61.1	74.3	61.1	75.2	79.2	83.3
Delaware	58.6	95.7	—	—	64.6	97.1
District of Columbia	—	—	—	—	71.9	76.5
Florida	95.3	96.8	88.4	97.7	95.8	98.2
Georgia	90.2	95.3	—	—	93.0	95.8
Hawaii	—	—	—	—	93.4	96.7
Idaho	—	—	63.7	84.8	—	—
Illinois	81.0	93.8	87.7	89.7	83.9	89.8
Indiana	80.7	94.8	—	—	92.1	99.2
Iowa	—	—	—	—	—	—
Kansas	57.5	—	—	—	83.9	86.5
Kentucky	85.2	—	—	—	91.6	100.0
Louisiana	95.4	98.8	—	—	98.8	96.8
Maine	—	—	—	—	—	—
Maryland	—	—	—	—	—	—
Massachusetts	69.5	—	80.2	89.7	—	—
Michigan	—	—	—	—	—	—
Minnesota	—	—	—	—	81.9	—
Mississippi	93.1	93.8	—	—	92.1	88.7
Missouri	62.9	85.9	—	—	63.9	74.5
Montana	—	—	—	—	—	—
Nebraska	—	—	—	—	—	—
Nevada	—	—	93.9	97.3	83.7	87.4
New Hampshire	—	—	—	—	59.6	—
New Jersey	85.3	91.3	85.6	92.8	87.0	97.0
New Mexico	—	—	72.2	73.8	71.6	85.9
New York	77.2	83.5	84.7	85.6	65.0	70.2
North Carolina	98.0	97.8	—	—	99.5	97.4
North Dakota	—	—	—	—	—	—
Ohio	87.6	—	—	—	81.4	—
Oklahoma	96.6	93.2	—	—	—	—
Oregon	—	—	—	93.5	—	—
Pennsylvania	78.4	94.9	—	—	81.0	96.9
Rhode Island	—	—	89.5	97.4	—	—
South Carolina	91.2	84.9	—	—	93.6	88.6
South Dakota	—	—	—	—	73.3	78.5
Tennessee	96.6	86.6	—	—	97.7	81.1
Texas	88.7	96.1	96.8	97.4	—	—
Utah	—	—	—	—	—	—
Vermont	—	—	—	—	61.1	73.0
Virginia	87.1	96.9	—	—	—	—
Washington	—	—	70.1	—	—	—
West Virginia	—	—	—	—	—	—
Wisconsin	54.5	—	—	—	86.7	87.2
Wyoming	—	—	—	—	95.4	92.8

— Not available.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

STATE ASSESSMENT RESULTS FOR NAEP SAMPLES AND SUMMARY FIGURES REPORTED BY STATES

All of the comparisons in this report were based on NAEP and state assessment data for the same schools, weighted by NAEP sampling weights to represent the public school students in the state. Theoretically, the weighted average of the state assessment scores in NAEP schools is an unbiased estimate of state-level statistics. There are several explanations for discrepancies between official state figures and results based on aggregation of state assessment results in the NAEP schools. Suppression of scores in some schools due to small number of students, failure to match state assessment scores to some NAEP schools, inclusion of different categories of schools and students in state figures, and summarization of scores in state reports to facilitate communication, can distort state-level estimates from NAEP schools. Tables 20 and 21 show the percentages of students meeting the primary standard for NAEP samples and states' published reports of mathematics achievement, for grades 4 and 8 respectively.

There are several reasons for failure to match some NAEP schools. For example, in states in which the only results available to compare to NAEP grade 4 results are grade 3 statistics, there might be a few NAEP schools that serve only grades 4 to 6, and these would have no grade 3 state assessment scores. Similarly, in sampling, NAEP does not cover special situations such as home schooling, and these may be included in state statistics. Finally, in reporting, to be succinct a state may issue reports with single summaries of scores across grades, while the data we analyzed might be specifically grade 4 scores. In fact, because NAEP samples are drawn with great care, factors such as these are more likely sources of discrepancies in tables 20 and 21 than sampling variation.

Table 20. Percentages of grade 4 students meeting primary standard of mathematics achievement in NAEP samples and states' published reports, by state: 2000 and 2003

State/jurisdiction	NAEP		State reports	
	2000	2003	2000	2003
Alabama	—	—	—	—
Alaska	—	67.2	65.0	64.8
Arizona	40.6	49.8	35.0	49.0
Arkansas	35.6	60.1	37.0	60.0
California	52.3	—	51.0	—
Colorado	—	86	—	56.0
Connecticut	63.7	58.6	60.2	60.4
Delaware	—	72.7	62.0	71.0
District of Columbia	36.9	—	—	—
Florida	—	55.4	46.0	54.0
Georgia	61.6	73.9	62.0	74.0
Hawaii	64.8	67.4	—	—
Idaho	—	76.6	—	77.5
Illinois	52.7	68.3	57.3	68.3
Indiana	—	—	73.0	66.0
Iowa	—	75.2	71.0	75.0
Kansas	59.3	73.7	62.4	73.6
Kentucky	31.1	37.9	31.3	38.1
Louisiana	11.3	15.5	12.0	16.0
Maine	23.3	29.1	23.0	28.0
Maryland	45.6	—	—	55.0
Massachusetts	41	38	40.0	40.0
Michigan	76.9	—	74.8	65.0
Minnesota	47.9	58.2	45.6	57.0
Mississippi	—	74	—	73.7
Missouri	36.6	36.7	36.7	37.2
Montana	—	75.3	—	73.0
Nebraska	60	—	—	—
Nevada	—	50.8	—	—
New Hampshire	—	80.3	40.0	42.0
New Jersey	—	67.6	65.8	—
New Mexico	—	42	—	—
New York	67.8	78.8	65.0	79.0
North Carolina	84.6	92.2	—	92.1
North Dakota	—	59	—	—
Ohio	42.3	59	—	58.0
Oklahoma	85.8	69.3	85.0	72.0
Oregon	67.2	77.8	70.0	76.0
Pennsylvania	—	56.8	52.0	56.3
Rhode Island	20.7	41.8	—	42.6
South Carolina	22.9	32.6	24.0	33.7
South Dakota	—	72.5	—	—
Tennessee	—	—	—	—
Texas	88.5	—	87.0	87.0
Utah	—	—	—	—
Vermont	47	52.9	47.3	53.0
Virginia	—	—	63.0	—
Washington	—	54	41.8	55.2
West Virginia	—	—	—	—
Wisconsin	71.9	—	74.0	71.0
Wyoming	25.6	36.4	27.0	37.0

— Not available.

NOTE: Primary standard is the state's standard for *proficient* performance.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. State reports are from state education agency websites.

Table 21. Percentages of grade 8 students meeting primary standard of mathematics achievement in NAEP samples and states' published reports, by state: 2000 and 2003

State/jurisdiction	NAEP		State report	
	2000	2003	2000	2003
Alabama	—	—	—	—
Alaska	—	65.2	39.0	63.8
Arizona	18.2	20.6	18.0	21.0
Arkansas	13.3	21.9	14.0	22.0
California	48	—	48.0	—
Colorado	—	68.2	35.0	38.0
Connecticut	57.2	56.0	54.8	56.1
Delaware	—	48.2	41.0	47.0
District of Columbia	8.6	—	—	—
Florida	—	54.1	51.0	56.0
Georgia	54.8	66.4	54.0	67.0
Hawaii	60.5	54.1	—	—
Idaho	—	52.5	—	53.0
Illinois	45.8	53.6	46.8	53.1
Indiana	—	—	63.0	66.0
Iowa	—	71.7	—	73.6
Kansas	55.3	59.3	54.6	60.0
Kentucky	26.0	31.9	25.2	30.9
Louisiana	7.3	8.7	8.0	8.0
Maine	20.6	17.1	21.0	18.0
Maryland	51.1	—	—	39.7
Massachusetts	33.7	38.1	34.0	37.0
Michigan	—	—	—	52.0
Minnesota	—	—	—	—
Mississippi	—	46.0	—	48.1
Missouri	12.9	13.4	14.0	13.9
Montana	—	70.3	—	69.0
Nebraska	59.7	—	—	—
Nevada	—	—	—	—
New Hampshire	—	—	—	—
New Jersey	—	56.2	59.7	—
New Mexico	—	39.4	—	—
New York	41.4	54.0	40.0	51.0
North Carolina	80.7	82.2	—	82.4
North Dakota	—	43.6	—	—
Ohio	—	—	—	—
Oklahoma	70.5	71.3	71.0	73.0
Oregon	54.7	57.6	56.0	59.0
Pennsylvania	—	51.5	52.0	51.3
Rhode Island	26.5	35.3	—	35.2
South Carolina	19.3	20.2	20.0	19.2
South Dakota	—	57.5	—	—
Tennessee	—	—	—	—
Texas	89.8	—	91.0	72.0
Utah	—	—	—	—
Vermont	45.6	51.7	47.0	51.7
Virginia	—	—	71.0	—
Washington	—	36.4	28.2	36.8
West Virginia	—	—	—	—
Wisconsin	42.9	—	30.0	73.0
Wyoming	31.5	35.1	32.0	35.0

— Not available.

NOTE: Primary standard is the state's standard for *proficient* performance.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. State reports are from state education agency websites.

National Assessment of Educational Progress

References

Campbell, J.R., Hombo, C.M., and Mazzeo, J. (2000). NAEP 1999 *Trends in Academic Progress: Three Decades of Student Performance*. (NCES 2000-469). U.S. Department of Education, Washington, DC: National Center for Education Statistics.

Forgione Jr., Pascal D. (1999), *Issues surrounding the release of the 1998 NAEP reading report card*. Testimony to the Committee on Education and the Workforce, U.S. House of Representatives, on May 27, 1999. Retrieved May 3, 2007, from http://www.house.gov/ed_workforce/hearings/106th/oi/naep52799/forgione.htm

Holland, P.W. (2002). Two measures of change in the gaps between the cdf's of test score distributions. *Journal of Educational and Behavioral Statistics*, 27: 3-17.

Lehr, C., and Thurlow, M. (2003). *Putting it all together: Including students with disabilities in assessment and accountability systems* (Policy Directions No. 16). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/Policy16.htm

McLaughlin, D.H. (2000). *Protecting State NAEP Trends from Changes in SD/LEP Inclusion Rates* (Report to the National Institute of Statistical Sciences). Palo Alto, CA: American Institutes for Research.

McLaughlin, D.H. (2001). *Exclusions and accommodations affect State NAEP gain statistics: Mathematics, 1996 to 2000* (Appendix to chapter 4 in the NAEP Validity Studies Report on Research Priorities). Palo Alto, CA: American Institutes for Research.

McLaughlin, D.H. (2003). *Full-Population estimates of reading gains between 1998 and 2002* (Report to NCES supporting inclusion of full population estimates in the report of the 2002 NAEP reading assessment). Palo Alto, CA: American Institutes for Research.

McLaughlin, D. (2005, December 8-9). *Considerations in using the longitudinal school-level state assessment score database*. Paper submitted to The National Academies Symposium on the Use of School-Level Data for Evaluating Federal Education Programs.

McLaughlin, D., Bandeira de Mello, V., Blankenship, C., Chaney, K., Esra, P., Hikawa, H., Rojas, D., William, P., and Wolman, M. (2008). *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.

National Assessment Governing Board–NAGB (2001). *Using the National Assessment of Educational Progress To Confirm State Test Results*, A Report of The Ad Hoc Committee on Confirming Test Results, March 2001.

National Center for Education Statistics (2005). *NAEP 2004 Trends in Academic Progress: Three Decades of Student Performance in Reading and Mathematics: Findings in Brief* (NCES 2005–463). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC: Government Printing Office

Ravitch, D. (2004). *A Brief History of Testing and Accountability*, Hoover Digest, 2004. Retrieved from the World Wide Web:
http://www.hoover.org/publications/digest/4495866.html

Shkolnik, J., and Blankenship, C. (2006), *School-level achievement data: A comparison of the use of mean scale scores and proficiency rates*, Palo Alto, CA: American Institutes for Research.

U.S. Department of Education (2002), *No Child Left Behind: A Desktop Reference*, Washington, DC: Office of Elementary and Secondary Education.

Vinovskis, M. A.(1998). *Overseeing the Nation's Report Card: The creation and evolution of the National Assessment Governing Board (NAGB)*, Paper prepared for the National Assessment Governing Board (NAGB), November 1998. Retrieved from the World Wide Web http://www.nagb.org/pubs/95222.pdf

Wise, L.L., Le, H., Hoffman, R.G., & Becker, D.E. (2004). *Testing NAEP full population estimates for sensitivity to violation of assumptions*. Technical Report TR-04-50. Alexandria, VA: HumRRO.

Wise, L.L., Le, H., Hoffman, R.G., & Becker, D.E. (2006). *Testing NAEP full population estimates for sensitivity to violation of assumptions*: Phase II Draft Technical Report DTR-06-08. Alexandria, VA: HumRRO.

A Methodological Notes

This appendix describes in more detail some of the methods used in this report. It covers technical aspects of (a) the placement of state achievement standards on the NAEP scale, and (b) the construction of a population achievement profiles based on school level averages. Finally, the estimation of the achievement of NAEP excluded students is discussed briefly; further details regarding this methodology as applied to reading assessments can be found in McLaughlin (2003).

ESTIMATING THE PLACEMENT OF STATE ACHIEVEMENT STANDARDS ON THE NAEP SCALE

If an achievement standard can be operationalized as a cutpoint on the NAEP scale, it is straightforward to estimate the percentage of the students in a state who meet that standard from the NAEP data. One compares each *plausible value* on the achievement scale assigned to a NAEP student, based on his or her responses to the test items, to the cutpoint of the standard. If it is greater than the cutpoint, the student's *weight* (the number of students in the population he or she represents) is added to the count of those meeting the standard; otherwise it is added to the count of those not meeting the standard.

If we had both the NAEP data for a state and the percentage of students in the state who met a NAEP standard, sorting all the plausible values in ascending order and determining which one just corresponds to the percent meeting the standard would be a straightforward task. For example, if the percent meeting the standard were given as 25, we would count down from the top of the order, adding up the weights until we reached 25 percent of the total weight for the state. This would not be exact, because there is some space between every pair of plausible values in the database, but with typically more than 2,000 NAEP participants in a state, we would expect it to be very close.

In this equipercentile mapping method, the standard error is an estimate of how far our estimate can be wrong, on average. The standard error is clearly related to the number of NAEP participants in the state.

Next, suppose that the percent meeting the standard is for the state's own assessment of achievement, not for NAEP's standard. We could carry out the same procedure to

find an estimate of the NAEP scale value corresponding to the state's standard; that is, the *cutpoint* for the state standard. Again, its standard error would depend on how large the NAEP sample of students is.

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

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

b. estimate the state proportion of students meeting the standard on the state test by weighting the proportions (from *step a*) for the NAEP schools, using NAEP weights;

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

d. find the point on the NAEP scale at which the estimated proportion of students in the state scoring above that point (using the distribution obtained in *step c*) equals the proportion of students in the state meeting the state's own performance standard (obtained in *step b*).

Operationally, the reported percentage meeting the state's standard in each NAEP school s, $p_s^{[STATE]}$, is used to compute a state percentage meeting the state's standards, using the NAEP school weights, w_s. For each school, w_s is the sum of the student weights, w_{is}, for the students selected for NAEP in that school.[1] For each of the five sets of NAEP plausible values, $v=1, 2, 3, 4, 5$, we solve the following equation for c, the point on the NAEP scale corresponding to the percentage meeting the state's standard:

$$p^{[STATE]} = \sum_{is} w_{is} p_s^{[STATE]} \Big/ \sum_{is} w_{is} = \sum_{is} w_{is} \partial_{isv}^{[NAEP]}(c) \Big/ \sum_{is} w_{is}$$

where the sum is over students in schools participating in NAEP, and $\partial_{isv}^{[NAEP]}(c)$ is equal to 1 if the v-th plausible value for student i in school s, y_{isv}, is greater than or equal to c. 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.

Specifically, each of the five parallel sets of NAEP plausible values (in the combined set of NAEP schools with matching state data) is sorted in increasing order. Then, for each plausible value in set v, y_v, the percentage of the NAEP student distribution that is greater than or equal to y_v, $p_v^{[NAEP]}(y_v)$, is computed. The two values of $p_v^{[NAEP]}(y_v)$ closest to $p^{[STATE]}$ above and below are identified, $p_{Uv}^{[NAEP]}(y_{Uv})$ and $p_{Lv}^{[NAEP]}(y_{Lv})$; and a point solution for c_v is identified by linear interpolation between y_{Uv} and y_{Lv}.

----

1. To ensure that NAEP and state assessments are fairly matched, NAEP schools which are missing state assessment scores (i.e., small schools, typically representing approximately four percent of the students in a state) are excluded from this process. Even if the small excluded schools are higher or lower performing than included schools, that should introduce no substantial bias in the estimation process, unless their high or lower scoring was specific to NAEP or specific to the state assessment.

Variation in results over the five sets of plausible values is a component of the standard error of the estimate, and the average of the five results is the reported mapping of the standard onto the NAEP scale.

The problem with this simple method is that it could be applied to any percentage, not just the percent meeting the state's achievement standard. Finding which NAEP scale score corresponds to the percent of the students, say, *living in large cities* would yield a value that is meaningless. A method is needed for testing the assumption that the percent we are given really corresponds to achievement on the NAEP scale.

The method we use to test the assumption is based on the fact that we are given the percent meeting the standard for each school participating in NAEP (and NAEP tests a random, representative sample of grade 4 or grade 8 students in each participating school). If the percentage we obtain from the state assessment for each school corresponds to achievement on the NAEP scale in that school, then applying the cutpoint, c, we estimated for the whole state to the NAEP plausible values in that school should yield an estimate of the percent meeting the standard in that school, based on NAEP, which matches the reported percent from the state assessment:

$$p_s^{[STATE]} = p_s^{[NAEP]}(c).$$

Of course, our estimated percentage meeting the state standard will not be exactly the same as the reported percent, because (a) NAEP only tests 20 to 25 students in the school, and (b) tests are not perfectly reliable. Moreover, in some states, we are given a grade 5 score for the state standard; in such cases, for the mapping method to be valid, we must assume that, on average across the state, the same percent of fourth graders would meet a grade 4 achievement standard as the percent of fifth graders who met the grade 5 standard. Of course, that would mean that our estimate for each school would have greater error—not only do some students learn more between fourth grade and fifth grade than others, but each cohort of students is different from the preceding grade's, in both random and systematic ways.

We need to have an estimate of how much error to expect in guessing each school's percent meeting the state's standard, to which we can compare the actual error. For this, we estimate the size of error if the state's standard were actually parallel to NAEP, and check whether the observed error is more similar to that or to the size of error we would expect if the reported percentages for the schools were unrelated to NAEP. If the error is sufficiently large, that calls the accuracy of the estimated standard into question.

Test criterion for the validity of the method

The test criterion is based on an evaluation of the discrepancies between (a) individual schools' reported percentages of students meeting a state standard and (b) the percentages of the NAEP achievement distribution that are greater than the NAEP cutpoint estimated for the state to be equivalent to that state standard. The method of estimation ensures that, on the average, these percentages agree, but there

is no assurance that they match for each school. To the extent that NAEP and the state assessment are parallel assessments, the percentages should agree for each school, but if NAEP and the state assessment are not correlated, then the linkage will not be able to reproduce the individual school results.

Failure of school percentages to match may also be due to student sampling variation, so the matching criterion must be a comparison of the extent of mismatch with the expectation based on random variation. To derive a heuristic criterion, we assume linear models for the school's reported percentage meeting standards on the state test, p_s, and the corresponding estimated percentage for school s, $\hat{p}_s$. The estimated percentage $\hat{p}_s$ is obtained by applying the linkage to the NAEP plausible values in the school.

$$p_s = \pi + \lambda_s + \delta_s + \gamma_s$$
$$\hat{p}_s = \pi + \lambda'_s + \delta'_s + \gamma_s$$

where π is the overall mean, λ and λ' are separate true between-school variations unique to the two assessments, δ and δ' are random sampling variations (due to the finiteness of the sample in each school), and γ is the common between-school variation between the reported and estimated percentages. We hope γ is large and λ and λ' are small. If λ and λ' are zero then the school-level scores would be reproducible perfectly except for random sampling and measurement error. Although linear models are clearly an oversimplification, they provide a way of distinguishing mappings of standards that are based on credible linkages from other mappings.

In terms of this model, the critical question for the validity of a mapping is whether the variation in γ, the achievement common to both assessments, as measured by $\sigma^2(\gamma)$, is large relative to variation in λ and λ', the achievement that is different between the assessments, as measured by $\sigma^2(\lambda)$ and $\sigma^2(\lambda')$. Because the linkage is constructed to match the variances of p_s and $\hat{p}_s$, the size of the measurement error variance, $\sigma^2(\delta)$ and $\sigma^2(\delta')$, does not affect the validity of the standard mapping, although it would affect the reproducibility of the school-level percentages.

The relative error criterion is the ratio of variance estimates:

$$k = [\sigma^2(\gamma) + (\sigma^2(\lambda) + \sigma^2(\lambda'))/2]/\sigma^2(\gamma)$$

The value of k is equal to 1 when there is no unique variance and to 2 when the common variance is the same as the average of the unique variances. Values larger than 1.5 indicate that the average unique variation is more than half as large as the common variation, which raises concern about the validity of the mapping.

We can estimate variance components from the variances of the sum and difference between the observed and estimated school-level percentages meeting a standard.

$$\sigma^2(p + \hat{p}) = \sigma^2(\lambda) + \sigma^2(\delta) + \sigma^2(\lambda') + \sigma^2(\delta') + 4\sigma^2(\gamma), \text{ and}$$

$$\sigma^2(p - \hat{p}) = \sigma^2(\lambda) + \sigma^2(\delta) + \sigma^2(\lambda') + \sigma^2(\delta').$$

By subtraction,

$$\sigma^2(\gamma) = (\sigma^2(p + \hat{p}) - \sigma^2(p - \hat{p}))/4.$$

To estimate $\sigma^2(\lambda)$ and $\sigma^2(\lambda')$, we compute $\sigma^2(p - \hat{p})$ for a case in which we know that there is no unique variation: using two of the five NAEP plausible value sets for the observed percentages and the other three for the estimated percentages. In this case,

$$\sigma^2(p_{NAEP} - \hat{p}_{NAEP}) = \sigma^2(\delta) + \sigma^2(\delta').$$

Substituting this in the equation for the variance of the differences, and rearranging terms,

$$\sigma^2(\lambda) + \sigma^2(\lambda') = \sigma^2(p - \hat{p}) - \sigma^2(p_{NAEP} - \hat{p}_{NAEP})$$

Substituting these estimates into the equation for k, we have

$$k = [\sigma^2(p + \hat{p}) - \sigma^2(p - \hat{p}) + 2\sigma^2(p - \hat{p}) - 2\sigma^2(p_{NAEP} - \hat{p}_{NAEP})] / (\sigma^2(p + \hat{p}) - \sigma^2(p - \hat{p}))$$

$$k = 1 + 2 [\sigma^2(p - \hat{p}) - \sigma^2(p_{NAEP} - \hat{p}_{NAEP}) / (\sigma^2(p + \hat{p}) - \sigma^2(p - \hat{p}))$$

That is, k is greater than 1 to the extent that the differences between observed and estimated school-level percentages are greater than the differences would be if both assessments were NAEP.[2]

The median values of k for primary grades 4 and 8 mathematics standards across states in 2003 were 1.241 and 1.171, corresponding to median values of 0.115 and 0.090 for $\sigma(p - \hat{p})$, 0.041 and 0.040 for $\sigma(p_{NAEP} - \hat{p}_{NAEP})$, and 0.216 and 0.197 for

2. The fact that the simulations are based on subsets of the NAEP data might lead to a slight over-estimate of $\sigma^2(\delta) + \sigma^2(\delta')$ because the distributions would not allow as fine-grained estimates of percentages achieving the standard. That over-estimate of random error in the linkage would, in turn, slightly reduce the estimate of k. In future work, one alternative to eliminate that effect might be to create a *parallel* NAEP by randomly imputing five additional plausible values for each participant, based on the mean and standard deviation of the five original plausible values. The result might increase the relative error measures slightly because they might reduce the term subtracted from the numerator of the formula for k.

$\sigma(p + \hat{p})/\sqrt{2}$, when p is measured on a [0,1] scale. A value of 1.5 for k corresponds to a common variance equal to the sum of the unique reliable variances in the observed and estimated percentages meeting the standards.

Setting the criterion for the validity of this application of the equipercentile mapping method at $k = 1.5$ (signifying equal amounts of common and unique variation) is arbitrary but plausible. Clearly, it should not be taken as an absolute inference of validity—two assessments, one with a relative criterion ratio of 1.6 and the other with 1.4, have similar validity. Setting a criterion serves to call attention to the cases in which one should consider a limitation on the validity of the mapping as an explanation for otherwise unexplainable results. While estimates of standards with greater relative error due to 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, irrespective of the relative error of the mapping of the standards.

Notes

With the relative error criterion we assessed the extent to which the error of the estimate is larger than it would be if NAEP and the state assessment were testing exactly the same underlying trait; in other words, by evaluating the accuracy with which each school's reported percentage of students meeting a state standard can be reproduced by applying the linkage to NAEP performance in that school. The method discussed here ensures that, on average, these percentages match, but there is no assurance that they match for each school. To the extent that NAEP and the state assessment are parallel assessments, the percentages should agree for each school, but if NAEP and the state assessment are not correlated, then the mapping will not be able to reproduce the individual school results. One difficult step in the validation process was estimating the amount of error to expect in reproducing the state-reported percentages for schools that could be due to random measurement and sampling error and not due to differences in the underlying traits being measured. For this purpose, we estimated the amount of error that would exist if both tests were NAEP. We used the distribution based on two plausible value sets to simulate the observed percent achieving the standard in each school and the distribution of the other three plausible value sets to simulate the estimated percent achieving the standard in the same school. The standard was the NAEP value determined (based on the entire state's NAEP sample) to provide the best estimate of the state's standard. Given the standard (a value on the NAEP scale), the percents achieving the standard are computed solely from the distribution of plausible values in the school. As an example, suppose the estimated standard is 225. For a school with 25 NAEP participants, there would be a distribution of 50 plausible values (two for each student) in the school for the simulated observed percent and 75 (three for each student) for the simulated estimated percent. The 50 plausible values for a school

represent random draws from the population of students in the state who (a) might have been in (similar) schools selected to participate in NAEP and (b) might have been selected to respond to one of the booklets of NAEP items. That distribution should be the same, except for random error, whether it is based on two, three, or five sets of plausible values.

CONSTRUCTING A POPULATION ACHIEVEMENT PROFILE BASED ON SCHOOL-LEVEL AVERAGES

For this report, individual scores on state assessments were not available. The comparisons in the report are based on school-level state assessment statistics and corresponding school-level summary statistics for NAEP. These school-level statistics include demographic breakdowns; that is, summary statistics for Black students in each school, Hispanic students in each school, and students eligible for free/reduced price lunch in each school. These are used in comparing NAEP and state assessment measurement of achievement gaps.

As defined in this report, a *population profile* of achievement is a percentile graph showing the distribution of achievement, from the lowest-performing students in a group at the left to the highest-performing students at the right. Concretely, one can imagine the students in a state lined up along the *x*-axis (i.e., the horizontal axis of a graph), sorted from left to right in order of increasing achievement scores, with each student's achievement score marked above him/her.[3]

When achievement scores are only available as school averages, or school averages for a particular category of students, the procedure, and the interpretation, is slightly different. Imagine the state's students in a demographic group lined up along the *x*-axis, sorted in order of average achievement of their group in their school (e.g., the percentage of students in the group who meet an achievement standard). Students in a school would be clustered together with others of their group in the same school. Each school's *width* on the *x*-axis would represent the *weight* of its students in representing the state population for the group. Thus, a school with many students in the demographic group would take up more space on the *x*-axis.

The population profile would then refer to the average performance in a school for the particular demographic group. The interpretation is similar to population profiles based on individual data, but there are fewer extremely high and extremely low scores. Gaps have the same average size because the achievement of each member of each demographic group is represented equally in the individually based and the school-level based profiles.[4] It is important to note that when we refer to school-level data, we are referring to aggregate achievement statistics for separate demographic groups in each school.

Because each school is weighted by the number of students in the demographic group it represents, we can still picture the population achievement profile as lining up the students in a state. They would be sorted from left to right in increasing order of their

---

3. See figure 1 in the text for an example of a population profile.

4. The exception to this is that due to suppression of small sample data in state assessment publications. As a result, students in schools with very small representations of a demographic group are underrepresented in school-level aggregates.

school's average achievement for their demographic group, with that average marked above them.

The procedure for computing a population profile based on school-level data can be described mathematically as follows. We start with a set of schools, j, in state i with N_i schools, with average achievement y_{igj}, for group g, whose weight in computing a state average for group g is w_{igj}. Sorting the schools on y_{igj}, creates a new subscript for schools, k, such that $y_{igk} \leq y_{ig(k+1)}$.

The sequence of histograms of height y_{igk}, width w_{igk}, for $k=1,\ldots, N_i$, forms a continuous distribution, which can be partitioned into one hundred equal intervals, $c = 1,\ldots, 100$, or percentiles. The achievement measure y_{igc}, for interval c in demographic group g in state i, is given by

$$y_{igc} = \left(\sum_{k = A_{igc}}^{B_{igc}} \overline{w}_{igk} \cdot y_{igk} \right) \bigg/ \left(\sum_{k = A_{igc}}^{B_{igc}} \overline{w}_{igk} \right)$$

where the A_{igc}, B_{igc}, and $\overline{w}_{igk}$ values are defined as follows:

$$\sum_{k = 1}^{A_{igc} - 1} w_{igk} \leq \frac{(c - 1)W}{100} \leq \sum_{k = 1}^{A_{igc}} w_{igk} \quad \text{and}$$

$$\sum_{k = 1}^{B_{igc} - 1} w_{igk} \leq \frac{(c)W}{100} \leq \sum_{k = 1}^{B_{igc}} w_{igk}$$

where W is the total weight.

For $A_{igc} + 1 \leq k \leq B_{igc} - 1$,

$$\overline{w}_{igk} = w_{igk};$$

$$\overline{w}_{A_{igc}} = \left(\sum_{k = 1}^{A_{igc}} w_{igk} - \frac{(c - 1)W}{100} \right), \text{ and}$$

$$\overline{w}_{B_{igk}} = \frac{(c)W}{100} - \left(\sum_{k = 1}^{B_{igc} - 1} w_{igk} \cdot w_{B_{igc}} \right).$$

It is important to note that the ordering of the schools for one group (e.g., Black students) may not be the same as the ordering of the same schools for another group (e.g., White students). Therefore, the gap between two population achievement profiles is not merely the within-school achievement gap; it combines both within-school and between-school achievement gaps to produce an overall achievement gap.

Standard errors can be computed for the individual y_{igc} by standard NAEP estimation methodology, computing a profile for each set of plausible values and for each set of replicate weights. However, in this report, we combined the percentiles into six groupings: the lowest and highest quartiles, the middle 50 percent, the lower and upper halves, and the entire range. The comparison of achievement between two groups for the entire range is mathematically equivalent to the average gap in the selected achievement measure.

ESTIMATING THE ACHIEVEMENT OF NAEP EXCLUDED STUDENTS

Since 1998, there has been concern that increasing or decreasing rates of exclusion of NAEP students from the sample might affect the sizes of gains reported by NAEP (e.g., Forgione, 1999; McLaughlin 2000, 2001, 2003). A method for imputing plausible values for the achievement of the excluded students based on ratings of their achievement by school staff has been applied to produce *full-population estimates*. The following description of the method is excerpted from a report to NCES on the application of the method to the estimation of reading gains between 1998 and 2002 (McLaughlin, 2003). The same method was used to produce full population estimates for mathematics in 2000 and 2003.

The method is made possible by the NAEP SD/LEP questionnaire, a descriptive survey filled out for each student with disability or English language learner selected to participate in NAEP - whether or not the student actually participates in NAEP or is excluded on the grounds that NAEP testing would be inappropriate for the student. The basic assumption of the method is that excluded students in each state with a particular profile of demographic characteristics and information on the SD/LEP questionnaire would, on average, be at the same achievement level as students with disabilities and English language learners who participated in NAEP in that state and had the same demographics and the same SD/LEP questionnaire profile.

The method for computing full-population estimates is straightforward. Plausible values are estimated for each excluded student, and these values are used along with the plausible values of included students to obtain state-level statistics. Estimation of plausible values for achievement of excluded students consists of two steps.

- Finding the combination of information on the teacher assessment form and demographic information that best accounts for variation in achievement of included students with disabilities and English language learners. At the same time, estimate the amount of error in that prediction.

- Combining the information in the same manner for excluded students to obtain a mean estimate for each student profile. Generate (five) plausible values for each student profile by adding to the mean estimate a random variable with the appropriate level of variation.

This method can be used to generate full-population estimates, either using the score on accommodated tests or not. It would work equally well after setting accommodated scores to *missing*. Because NAEP is currently using accommodated scores, the full-population estimates presented here treat accommodated scores as valid indicators of achievement. The procedure was carried out separately for each grade and subject each year.

In 2004, the NAEP Quality Assurance contractor, the Human Resources Research Organization (HumRRO), tested the methodology used in this report to estimate the performance of the excluded students for sensitivity to violation of assumptions

(Wise et al., 2004). Overall, under the assumptions of the model, the full population estimates were unbiased. Violations of these assumptions led to slightly biased estimates which, at the jurisdiction level, were considered negligible.

The Education Testing Service (ETS) has recently developed an alternative approach to address the exclusion problem. ETS's approach is also an imputation procedure; it is based on the same basic assumptions used by AIR, with the only substantive difference being the inclusion of the school's state assessment score variable in the imputation model.[5] When both approaches were compared (Wise et al., 2006), their performances were equivalent. When model assumptions were violated, the ETS estimates were slightly less biased but, overall, the two approaches produced similar standard error estimates (Wise et al., 2006).[6]

The overall conclusion is that "the degree of bias in mean estimates generated from the FPE method was quite small and represented a significant improvement over simply ignoring excluded students even when excluded students' achievement levels were much lower than predicted by background information."

---

5. AIR deliberately excluded that variable in order to eliminate the argument that NAEP's FPE imputation might be based on something other than NAEP. For example, using state assessment results that include accommodations not allowed by NAEP may negatively impact the credibility of NAEP estimates for the excluded students.

6. The small differences between the two models seem to be mostly related to the inclusion of school's state assessment score variable in the ETS model.

B Tables of State Mathematics Standards and Tables of Standard Errors

Table B-1. NAEP equivalent of state grade 4 mathematics achievement standards, by state: 2003

State/jurisdiction	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5
Alabama	—	—	—	—	—
Alaska	—	—	222.6	—	—
Arizona	—	191.6	230.4	239.5	—
Arkansas	—	208.4	222.8	238.4	—
California	178.9	208.4	230.1	254.6	—
Colorado	—	—	200.4	231.8	260.6
Connecticut	202.1	215.2	234.9	262.9	—
Delaware	—	198.3	217.0	257.6	271.1
District of Columbia	—	190.8	218.0	249.3	—
Florida	—	210.5	229.9	257.8	280.9
Georgia	—	—	211.1	256.3	—
Hawaii	—	205.1	251.3	283.0	—
Idaho	—	183.3	216.1	253.9	—
Illinois	—	173.2	216.3	271.8	—
Indiana	—	—	225.1	270.1	—
Iowa	—	—	220.4	—	—
Kansas	—	202.3	225.2	241.5	263.2
Kentucky	—	216.4	236.9	264.8	—
Louisiana	201.5	220.7	254.8	276.6	—
Maine	—	222.4	251.9	281.2	—
Maryland	—	—	227.7	274.0	—
Massachusetts	—	213.7	250.7	274.1	—
Michigan	—	190.0	225.1	258.8	—
Minnesota	192.6	219.5	236.5	267.2	—
Mississippi	—	185.2	203.4	—	—
Missouri	180.1	211.7	243.8	272.9	—
Montana	—	204.0	219.0	260.7	—
Nebraska	—	—	—	—	—
Nevada	—	206.1	227.1	248.0	—
New Hampshire	—	—	220.3	247.9	269.1
New Jersey	—	—	226.4	259.3	—
New Mexico	—	205.7	228.3	246.9	—
New York	—	184.8	211.8	249.6	—
North Carolina	—	184.4	201.4	235.7	—
North Dakota	—	—	233.1	—	—
Ohio	—	221.1	231.2	264.5	—
Oklahoma	—	171.5	216.1	255.9	—
Oregon	—	—	213.9	254.3	—
Pennsylvania	—	211.6	232.1	254.1	—
Rhode Island	—	—	237.2	—	—
South Carolina	—	211.0	247.3	265.4	—
South Dakota	—	168.8	222.8	258.4	—
Tennessee	—	—	—	—	—
Texas	—	—	203.3	—	—
Utah	—	—	—	—	—
Vermont	181.4	224.7	239.9	268.0	—
Virginia	—	—	219.7	264.0	—
Washington	—	213.6	235.1	252.6	—
West Virginia	—	—	—	—	—
Wisconsin	—	210.1	221.6	252.6	—
Wyoming	—	228.5	250.0	273.0	—

— Not available.

NOTE: Standard 3 represents the primary standard. In most cases, it is the criterion for Adequate Yearly Progress (AYP).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-2. Standard errors for table B-1: NAEP equivalent of state grade 4 mathematics achievement standards, by state: 2003

State/ jurisdiction	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5
Alabama	†	†	†	†	†
Alaska	†	†	0.80	†	†
Arizona	†	1.72	0.81	0.94	†
Arkansas	†	1.37	0.88	1.09	†
California	1.40	1.19	1.42	1.61	†
Colorado	†	†	1.22	0.89	0.91
Connecticut	1.97	0.97	0.83	1.05	†
Delaware	†	1.80	1.00	1.66	1.36
District of Columbia	†	1.35	0.99	1.32	†
Florida	†	1.04	0.70	0.86	1.46
Georgia	†	†	0.64	1.03	†
Hawaii	†	0.97	0.82	2.64	†
Idaho	†	1.95	0.93	1.02	†
Illinois	†	2.11	0.85	1.11	†
Indiana	†	†	1.12	1.06	†
Iowa	†	†	0.94	†	†
Kansas	†	1.33	1.24	0.85	1.02
Kentucky	†	1.12	1.20	1.19	†
Louisiana	0.78	0.92	1.26	1.17	†
Maine	†	1.29	0.71	1.11	†
Maryland	†	†	1.30	1.18	†
Massachusetts	†	1.42	0.78	0.97	†
Michigan	†	1.81	0.98	0.83	†
Minnesota	1.61	1.21	0.81	1.00	†
Mississippi	†	1.73	0.87	†	†
Missouri	1.91	1.47	0.86	1.33	†
Montana	†	1.12	0.82	0.71	†
Nebraska	†	†	†	†	†
Nevada	†	1.27	0.86	0.63	†
New Hampshire	†	†	1.51	1.13	0.86
New Jersey	†	†	1.18	0.93	†
New Mexico	†	1.66	1.16	1.89	†
New York	†	2.42	1.21	1.00	†
North Carolina	†	2.44	2.16	0.67	†
North Dakota	†	†	0.96	†	†
Ohio	†	0.72	0.86	1.35	†
Oklahoma	†	2.20	0.90	1.05	†
Oregon	†	†	1.66	0.70	†
Pennsylvania	†	1.42	1.26	1.00	†
Rhode Island	†	†	0.66	†	†
South Carolina	†	1.27	0.91	1.15	†
South Dakota	†	3.96	0.64	1.01	†
Tennessee	†	†	†	†	†
Texas	†	†	1.43	†	†
Utah	†	†	†	†	†
Vermont	1.92	0.88	0.98	1.37	†
Virginia	†	†	1.30	1.27	†
Washington	†	1.01	1.35	1.11	†
West Virginia	†	†	†	†	†
Wisconsin	†	0.98	0.93	0.54	†
Wyoming	†	0.63	0.52	0.72	†

† Not applicable.

NOTE: Standard 3 represents the primary standard. In most cases, it is the criterion for Adequate Yearly Progress (AYP).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-3. NAEP equivalent of state grade 8 mathematics achievement standards, by state: 2003

State/ jurisdiction	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5
Alabama	—	—	—	—	—
Alaska	—	—	267.7	—	—
Arizona	—	262.0	299.3	321.2	—
Arkansas	—	250.6	295.3	332.2	—
California	214.8	252.1	286.3	322.5	—
Colorado	—	—	266.9	294.0	322.1
Connecticut	231.4	255.7	277.9	312.2	—
Delaware	—	250.3	275.2	304.9	318.4
District of Columbia	—	250.2	282.5	322.2	—
Florida	—	243.9	267.7	299.0	319.4
Georgia	—	—	254.2	307.3	—
Hawaii	—	244.4	298.5	345.2	—
Idaho	—	246.8	279.8	315.7	—
Illinois	—	214.4	274.0	313.1	—
Indiana	—	—	268.4	321.9	—
Iowa	—	—	265.4	—	—
Kansas	—	253.9	277.2	295.7	316.6
Kentucky	—	251.7	290.1	320.1	—
Louisiana	242.6	264.1	310.2	325.6	—
Maine	—	267.6	310.5	358.0	—
Maryland	—	—	284.7	317.8	—
Massachusetts	—	269.3	298.2	325.9	—
Michigan	—	251.8	276.8	297.1	—
Minnesota	—	—	—	—	—
Mississippi	—	239.1	260.0	—	—
Missouri	244.8	280.1	313.8	347.2	—
Montana	—	252.8	270.1	318.1	—
Nebraska	—	—	—	—	—
Nevada	—	250.6	276.4	299.5	—
New Hampshire	—	—	—	—	—
New Jersey	—	—	277.8	317.2	—
New Mexico	—	241.4	269.2	294.3	—
New York	—	238.7	277.2	323.5	—
North Carolina	—	217.1	244.4	283.5	—
North Dakota	—	—	292.7	—	—
Ohio	—	—	—	—	—
Oklahoma	—	212.2	254.8	304.1	—
Oregon	—	—	274.1	297.9	—
Pennsylvania	—	255.6	278.8	309.5	—
Rhode Island	—	—	286.1	—	—
South Carolina	—	259.9	304.7	326.2	—
South Dakota	—	220.9	281.0	328.8	—
Tennessee	—	—	—	—	—
Texas	—	—	255.4	—	—
Utah	—	—	—	—	—
Vermont	239.4	269.4	285.1	316.4	—
Virginia	—	—	252.2	312.5	—
Washington	—	277.1	294.5	311.2	—
West Virginia	—	—	—	—	—
Wisconsin	—	239.4	259.2	305.2	—
Wyoming	—	267.3	296.5	320.9	—

— Not available.

NOTE: Standard 3 represents the primary standard. In most cases, it is the criterion for Adequate Yearly Progress (AYP).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

National Assessment of Educational Progress

Table B-4. Standard errors for table B-3: NAEP equivalent of state grade 8 mathematics achievement standards, by state: 2003

State/ jurisdiction	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5
Alabama	†	†	†	†	†
Alaska	†	†	1.29	†	†
Arizona	†	1.54	1.12	1.37	†
Arkansas	†	1.56	1.34	2.33	†
California	2.04	1.92	1.35	1.43	†
Colorado	†	†	0.92	0.79	1.19
Connecticut	1.95	1.37	0.89	1.13	†
Delaware	†	0.97	0.73	1.32	1.22
District of Columbia	†	0.78	1.11	1.55	†
Florida	†	1.24	1.24	1.32	1.37
Georgia	†	†	1.24	2.62	†
Hawaii	†	1.41	1.86	3.80	†
Idaho	†	1.28	0.73	0.91	†
Illinois	†	1.29	1.56	1.39	†
Indiana	†	†	1.25	1.28	†
Iowa	†	†	1.01	†	†
Kansas	†	1.66	0.88	1.14	1.08
Kentucky	†	0.87	0.87	1.41	†
Louisiana	1.53	0.97	1.56	2.73	†
Maine	†	1.07	0.77	4.01	†
Maryland	†	†	1.09	1.59	†
Massachusetts	†	1.31	0.67	0.80	†
Michigan	†	1.65	0.87	1.13	†
Minnesota	†	†	†	†	†
Mississippi	†	1.23	0.86	†	†
Missouri	1.64	0.59	1.36	2.08	†
Montana	†	1.64	0.93	1.09	†
Nebraska	†	†	†	†	†
Nevada	†	1.03	0.61	0.86	†
New Hampshire	†	†	†	†	†
New Jersey	†	†	0.96	1.03	†
New Mexico	†	1.36	1.59	1.03	†
New York	†	1.00	1.06	1.28	†
North Carolina	†	3.12	2.03	1.10	†
North Dakota	†	†	1.24	†	†
Ohio	†	†	†	†	†
Oklahoma	†	2.73	1.48	1.49	†
Oregon	†	†	1.17	1.70	†
Pennsylvania	†	0.81	1.13	0.80	†
Rhode Island	†	†	1.43	†	†
South Carolina	†	1.11	1.24	2.06	†
South Dakota	†	3.02	0.57	1.57	†
Tennessee	†	†	†	†	†
Texas	†	†	1.02	†	†
Utah	†	†	†	†	†
Vermont	1.59	0.94	1.10	1.37	†
Virginia	†	†	1.52	1.23	†
Washington	†	1.37	1.07	0.81	†
West Virginia	†	†	†	†	†
Wisconsin	†	1.28	1.54	0.78	†
Wyoming	†	1.28	1.22	0.85	†

† Not applicable.

NOTE: Standard 3 represents the primary standard. In most cases, it is the criterion for adequate yearly progress (AYP).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-5. Standard errors for table 4: Correlations between NAEP and state assessment school-level percentages meeting primary state mathematics standards, grades 4 and 8, by state: 2003

State/ jurisdiction	Name of standard	Grades for state assessment		Grade 4 standard error	Grade 8 standard error
Alabama	Percentile Rank	4	8	0.010	0.016
Alaska	Proficient	4	8	0.023	0.028
Arizona	Meeting	5	8	0.012	0.014
Arkansas	Proficient	4	8	0.009	0.025
California	Proficient	4	7	0.009	0.011
Colorado	Partially Proficient	5	8	0.013	0.017
Connecticut	Goal	4	8	0.004	0.007
Delaware	Meeting	5	8	0.035	0.041
District of Columbia	Proficient	4	8	0.003	0.008
Florida	(3) Partial Success	4	8	0.012	0.018
Georgia	Meeting	4	8	0.017	0.012
Hawaii	Meeting	5	8	0.010	0.017
Idaho	Proficient	4	8	0.039	0.026
Illinois	Meeting	5	8	0.011	0.009
Indiana	Pass	3	8	0.013	0.022
Iowa	Proficient	4	8	0.016	0.047
Kansas	Proficient	4	7	0.021	0.014
Kentucky	Proficient	5	8	0.019	0.026
Louisiana	Mastery	4	8	0.020	0.024
Maine	Meeting	4	8	0.052	0.036
Maryland	Proficient	5	8	0.003	0.016
Massachusetts	Proficient	4	8	0.008	0.012
Michigan	Meeting	4	8	0.011	0.009
Minnesota	(3) Proficient	5	—	0.016	†
Mississippi	Proficient	4	8	0.016	0.012
Missouri	Proficient	4	8	0.016	0.033
Montana	Proficient	4	8	0.020	0.033
Nebraska	Meeting	—	—	†	†
Nevada	Meeting:3	4	7	0.034	0.011
New Hampshire	Basic	—	—	0.017	†
New Jersey	Proficient	4	8	0.009	0.007
New Mexico	Top half	4	8	0.014	0.016
New York	Meeting	4	8	0.011	0.009
North Carolina	Consistent Mastery	4	8	0.044	0.016
North Dakota	Meeting	4	8	0.022	0.048
Ohio	Proficient	4	—	0.011	†
Oklahoma	Satisfactory	5	8	0.016	0.021
Oregon	Meeting	5	8	0.031	0.022
Pennsylvania	Proficient	5	8	0.008	0.011
Rhode Island	Proficient	4	8	0.011	0.014
South Carolina	Proficient	4	8	0.012	0.014
South Dakota	Proficient	4	8	0.011	0.008
Tennessee	Percentile Rank	4	8	0.016	0.027
Texas	Passing	4	8	0.052	0.009
Utah	Percentile Rank	5	8	0.008	0.013
Vermont	Achieved	4	8	0.021	0.026
Virginia	Proficient	5	8	0.017	0.028
Washington	Met	4	7	0.019	0.026
West Virginia	Top half	—	—	†	†
Wisconsin	Proficient	4	8	0.015	0.008
Wyoming	Proficient	4	8	0.018	0.023

— Not available.

† Not applicable.

NOTE: Primary standard is the state's standard for *proficient* performance.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-6. Standard errors for table 8: Mathematics achievement gains in percentage meeting the primary standard in grade 4, by state: 2000 and 2003

State/ jurisdiction	State			NAEP		
	2000	2003	Gain	2000	2003	Gain
Alabama	1.23	0.89	1.52	1.64	1.43	2.18
Alaska	†	1.23	†	†	1.45	†
Arizona	1.56	1.51	2.17	1.91	1.55	2.46
Arkansas	1.53	1.11	1.89	1.60	1.45	2.16
California	1.90	†	†	2.26	†	†
Colorado	†	0.94	†	†	1.06	†
Connecticut	1.18	0.98	1.53	1.75	1.19	2.12
Delaware	†	0.19	†	†	1.27	†
District of Columbia	0.43	†	†	1.18	†	†
Florida	†	1.08	†	†	1.48	†
Georgia	1.04	0.93	1.40	1.57	1.25	2.01
Hawaii	1.11	1.01	1.50	1.67	1.20	2.06
Idaho	†	1.18	†	†	1.06	†
Illinois	2.51	1.14	2.76	2.89	1.61	3.31
Indiana	†	†	†	†	†	†
Iowa	†	0.84	†	†	1.26	†
Kansas	1.87	1.29	2.27	2.22	1.47	2.66
Kentucky	1.15	1.42	1.83	1.81	1.62	2.43
Louisiana	0.81	0.92	1.23	1.17	1.18	1.66
Maine	1.13	0.77	1.37	1.47	1.42	2.04
Maryland	1.45	†	†	1.68	†	†
Massachusetts	1.23	0.94	1.55	1.57	1.34	2.06
Michigan	1.81	†	†	1.77	†	†
Minnesota	1.79	1.30	2.21	2.40	1.51	2.84
Mississippi	†	1.33	†	†	1.31	†
Missouri	1.44	0.94	1.72	1.78	1.55	2.36
Montana	†	1.02	†	†	1.12	†
Nebraska	4.05	†	†	6.90	†	†
Nevada	†	†	†	†	†	†
New Hampshire	†	0.89	†	†	1.10	†
New Jersey	†	1.31	†	†	1.68	†
New Mexico	†	1.38	†	†	1.86	†
New York	3.28	0.80	3.38	3.71	1.20	3.90
North Carolina	0.72	0.39	0.82	1.29	0.80	1.52
North Dakota	†	1.01	†	†	1.10	†
Ohio	2.09	1.51	2.58	2.38	1.72	2.94
Oklahoma	0.85	1.40	1.64	1.60	1.08	1.93
Oregon	1.68	1.06	1.99	2.51	1.18	2.77
Pennsylvania	†	1.83	†	†	1.85	†
Rhode Island	0.91	1.39	1.66	1.15	1.37	1.79
South Carolina	1.22	1.07	1.62	1.32	1.44	1.95
South Dakota	†	1.08	†	†	1.15	†
Tennessee	†	1.23	†	†	1.56	†
Texas	0.84	†	†	1.42	†	†
Utah	1.38	1.02	1.72	1.68	1.16	2.04
Vermont	1.52	0.90	1.77	2.57	1.18	2.83
Virginia	†	†	†	†	†	†
Washington	†	1.49	†	†	1.61	†
West Virginia	†	†	†	†	†	†
Wisconsin	1.34	†	†	1.99	†	†
Wyoming	1.38	0.29	1.41	1.74	1.09	2.05
Average gain	†	†	1.89	†	†	2.40

† Not applicable.

NOTE: Primary standard is the state's standard for *proficient* performance.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-7. Standard errors for table 9: Mathematics achievement gains in percentage meeting the primary standard in grade 8, by state: 2000 and 2003

State/jurisdiction	State			NAEP		
	2000	2003	Gain	2000	2003	Gain
Alabama	1.28	1.25	1.79	1.67	1.31	2.12
Alaska	†	0.67	†	†	1.45	†
Arizona	1.15	1.01	1.53	1.54	1.22	1.96
Arkansas	0.59	0.87	1.05	0.89	1.26	1.54
California	1.88	†	†	2.39	†	†
Colorado	†	0.96	†	†	1.25	†
Connecticut	1.26	1.04	1.63	1.61	1.51	2.21
Delaware	†	0.20	†	†	1.22	†
District of Columbia	0.45	†	†	1.30	†	†
Florida	†	1.46	†	†	1.63	†
Georgia	1.22	0.84	1.48	1.64	1.43	2.18
Hawaii	0.66	0.17	0.68	1.59	1.00	1.88
Idaho	†	0.56	†	†	1.15	†
Illinois	1.56	0.98	1.84	1.82	1.71	2.50
Indiana	†	†	†	†	†	†
Iowa	†	0.74	†	†	1.11	†
Kansas	1.43	1.25	1.90	2.07	1.34	2.47
Kentucky	1.15	1.19	1.65	1.58	1.43	2.13
Louisiana	0.73	0.65	0.98	0.94	1.06	1.42
Maine	0.69	0.44	0.82	1.37	0.89	1.63
Maryland	1.37	†	†	1.89	†	†
Massachusetts	1.05	0.80	1.32	1.45	1.00	1.76
Michigan	†	†	†	†	†	†
Minnesota	†	†	†	†	†	†
Mississippi	†	1.11	†	†	1.55	†
Missouri	0.60	0.61	0.86	0.93	1.08	1.43
Montana	†	0.78	†	†	1.10	†
Nebraska	7.11	†	†	14.83	†	†
Nevada	†	†	†	†	†	†
New Hampshire	†	†	†	†	†	†
New Jersey	†	1.16	†	†	1.45	†
New Mexico	†	0.76	†	†	1.52	†
New York	2.65	1.28	2.94	3.19	1.40	3.48
North Carolina	0.67	0.55	0.87	1.34	1.06	1.71
North Dakota	†	0.29	†	†	1.43	†
Ohio	†	†	†	†	†	†
Oklahoma	1.10	0.99	1.48	1.72	1.39	2.21
Oregon	1.26	0.88	1.54	2.14	1.58	2.66
Pennsylvania	†	0.97	†	†	1.56	†
Rhode Island	0.37	0.25	0.45	0.98	1.05	1.44
South Carolina	0.91	1.09	1.42	1.30	1.48	1.97
South Dakota	†	1.00	†	†	1.33	†
Tennessee	†	1.06	†	†	1.65	†
Texas	0.68	†	†	1.08	†	†
Utah	0.68	0.77	1.03	1.27	1.58	2.03
Vermont	0.89	0.71	1.14	1.59	1.05	1.91
Virginia	†	†	†	†	†	†
Washington	†	0.98	†	†	1.24	†
West Virginia	†	†	†	†	†	†
Wisconsin	1.28	†	†	1.98	†	†
Wyoming	0.61	0.16	0.63	1.19	0.95	1.52
Average gain	†	†	1.43	†	†	2.06

† Not applicable.

NOTE: Primary standard is the state's standard for *proficient* performance.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

National Assessment of Educational Progress

Table B-8. Standard errors for tables 10 and 11: Differences between NAEP and state assessments of grades 4 and 8 mathematics achievement race and poverty gaps, by state: 2003

State/ jurisdiction	Grade 4			Grade 8		
	Black-White	Hisp.-White	Poverty	Black-White	Hisp.-White	Poverty
Alabama	2.19	†	1.98	2.08	†	1.82
Alaska	†	†	†	†	†	†
Arizona	†	3.05	†	†	2.62	†
Arkansas	3.21	†	2.60	1.87	†	2.45
California	†	2.78	2.43	†	3.13	2.90
Colorado	†	†	†	†	†	†
Connecticut	3.65	3.90	2.71	3.82	3.18	2.74
Delaware	1.53	†	1.46	1.43	†	1.98
District of Columbia	†	†	3.81	†	†	1.97
Florida	2.22	2.39	1.94	2.16	2.61	1.91
Georgia	2.35	†	2.16	2.60	†	2.62
Hawaii	†	†	2.00	†	†	1.47
Idaho	†	4.19	†	†	2.87	†
Illinois	3.71	2.89	2.97	2.70	2.55	2.41
Indiana	4.86	†	2.88	4.52	†	2.57
Iowa	†	†	†	†	†	†
Kansas	5.30	†	4.32	†	†	2.98
Kentucky	2.78	†	2.48	†	†	2.12
Louisiana	2.02	†	2.58	1.59	†	1.68
Maine	†	†	†	†	†	†
Maryland	†	†	†	†	†	†
Massachusetts	2.54	3.80	†	†	2.90	†
Michigan	†	†	†	†	†	†
Minnesota	†	†	3.04	†	†	†
Mississippi	2.51	†	2.61	2.30	†	2.57
Missouri	3.17	†	4.24	1.49	†	1.94
Montana	†	†	†	†	†	†
Nebraska	†	†	†	†	†	†
Nevada	†	2.75	2.64	†	1.38	1.23
New Hampshire	†	†	3.97	†	†	†
New Jersey	4.20	4.09	4.82	2.94	3.32	2.82
New Mexico	†	3.87	3.88	†	2.44	3.22
New York	2.46	3.37	1.87	2.59	3.25	2.64
North Carolina	2.68	†	2.21	2.42	†	2.57
North Dakota	†	†	†	†	†	†
Ohio	2.68	†	2.35	†	†	†
Oklahoma	4.41	†	†	5.00	†	†
Oregon	†	†	†	†	3.73	†
Pennsylvania	4.22	†	3.70	2.95	†	2.91
Rhode Island	†	2.42	†	†	1.30	†
South Carolina	2.43	†	2.36	1.82	†	1.69
South Dakota	†	†	2.22	†	†	1.62
Tennessee	2.16	†	2.06	2.17	†	2.28
Texas	3.72	1.90	†	3.06	2.54	†
Utah	†	†	†	†	†	†
Vermont	†	†	2.51	†	†	2.44
Virginia	3.29	†	†	2.99	†	†
Washington	†	3.53	†	†	†	†
West Virginia	†	†	†	†	†	†
Wisconsin	4.56	†	2.38	†	†	3.15
Wyoming	†	†	1.20	†	†	1.19

† Not applicable.

NOTE: The poverty gap refers to the difference in achievement between economically disadvantaged students and other students, where disadvantaged students are defined as those eligible for free/reduced price lunch.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-9. Standard errors for tables in appendix D: Percentage of students identified both with a disability and as an English language learner, by state: 2000 and 2003

State/jurisdiction	Grade 4		Grade 8	
	2000	2003	2000	2003
Alabama	#	0.14	0.19	0.17
Alaska	†	0.61	†	0.44
Arizona	0.26	0.67	0.34	0.40
Arkansas	0.12	0.31	#	0.20
California	0.35	0.43	0.37	0.37
Colorado	†	0.27	†	0.21
Connecticut	0.13	0.20	0.17	0.18
Delaware	†	0.17	†	0.21
District of Columbia	0.11	0.20	0.13	0.25
Florida	†	0.39	†	0.32
Georgia	0.04	0.23	0.17	0.16
Hawaii	0.23	0.23	0.26	0.24
Idaho	0.28	0.23	0.35	0.22
Illinois	0.38	0.35	0.23	0.27
Indiana	0.16	0.18	0.07	0.30
Iowa	0.08	0.32	#	0.19
Kansas	0.39	0.13	#	0.23
Kentucky	0.09	0.23	0.16	0.14
Louisiana	0.19	0.28	#	0.21
Maine	0.13	0.27	0.15	0.20
Maryland	0.17	0.32	0.31	0.15
Massachusetts	0.20	0.22	0.26	0.20
Michigan	0.08	0.28	#	0.14
Minnesota	0.19	0.22	0.20	0.12
Mississippi	#	0.22	0.07	0.04
Missouri	0.07	0.18	#	0.12
Montana	#	0.58	0.21	0.25
Nebraska	0.12	0.30	0.18	0.16
Nevada	0.23	0.43	0.18	0.26
New Hampshire	†	0.19	†	0.11
New Jersey	†	0.26	†	0.18
New Mexico	0.71	0.67	0.72	0.60
New York	0.29	0.36	0.28	0.25
North Carolina	0.11	0.34	0.30	0.23
North Dakota	0.10	0.20	0.39	0.33
Ohio	#	0.28	0.59	0.13
Oklahoma	0.35	0.31	0.13	0.43
Oregon	0.22	0.32	0.27	0.26
Pennsylvania	†	0.24	†	0.34
Rhode Island	0.19	0.41	0.19	0.26
South Carolina	0.31	0.23	0.09	0.16
South Dakota	†	0.20	†	0.14
Tennessee	0.41	0.13	#	0.28
Texas	0.30	0.40	0.30	0.33
Utah	0.19	0.39	0.26	0.30
Vermont	0.08	0.17	0.11	0.17
Virginia	0.18	0.34	0.12	0.31
Washington	†	0.39	†	0.31
West Virginia	0.13	0.11	0.11	0.11
Wisconsin	0.27	0.22	0.16	0.19
Wyoming	0.46	0.21	0.42	0.25

† Not applicable.

Rounds to zero.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-10. Standard errors for tables in appendix D: Percentage of students identified with a disability, but not as an English language learner, by state: 2000 and 2003

State/jurisdiction	Grade 4		Grade 8	
	2000	2003	2000	2003
Alabama	0.81	0.59	0.86	0.82
Alaska	†	0.73	†	0.62
Arizona	0.90	0.59	0.92	0.69
Arkansas	0.84	0.79	0.90	0.87
California	0.82	0.46	0.91	0.54
Colorado	†	0.60	#	0.59
Connecticut	0.65	0.72	0.84	0.75
Delaware	†	0.59	†	0.72
District of Columbia	0.92	0.65	1.04	0.76
Florida	†	0.79	†	0.88
Georgia	0.71	0.58	0.69	0.55
Hawaii	0.66	0.57	1.43	0.60
Idaho	0.82	0.57	0.61	0.62
Illinois	0.82	0.70	0.93	0.68
Indiana	0.78	0.71	0.76	0.66
Iowa	1.06	0.78	#	0.77
Kansas	0.91	0.67	1.21	0.66
Kentucky	0.79	0.82	0.76	0.75
Louisiana	1.04	1.49	0.84	0.96
Maine	1.00	0.87	0.84	0.78
Maryland	0.73	0.74	0.83	0.94
Massachusetts	0.80	0.80	0.99	0.70
Michigan	0.89	0.69	0.80	0.69
Minnesota	0.84	0.63	1.15	0.58
Mississippi	0.53	0.54	0.73	0.51
Missouri	0.79	0.68	0.97	0.74
Montana	1.46	0.78	0.89	0.54
Nebraska	1.35	0.89	1.12	0.62
Nevada	0.67	0.71	0.90	0.69
New Hampshire	†	0.65	†	0.67
New Jersey	†	0.88	†	0.87
New Mexico	0.82	0.77	0.86	0.68
New York	0.97	0.61	0.73	0.70
North Carolina	0.80	0.70	0.81	0.81
North Dakota	0.95	0.79	0.90	0.63
Ohio	1.12	0.71	0.95	0.79
Oklahoma	0.81	0.86	0.91	0.79
Oregon	0.78	0.86	1.03	0.68
Pennsylvania	†	0.88	†	0.66
Rhode Island	0.84	0.96	1.31	0.60
South Carolina	0.83	0.72	0.74	0.83
South Dakota	†	0.67	†	0.55
Tennessee	0.94	0.80	0.87	0.90
Texas	0.75	0.71	0.73	0.76
Utah	0.57	0.59	0.80	0.59
Vermont	1.53	0.77	1.13	0.79
Virginia	0.70	0.72	0.71	0.68
Washington	†	0.73	†	0.64
West Virginia	0.81	0.89	0.81	0.88
Wisconsin	1.17	0.88	0.92	0.68
Wyoming	1.09	0.62	1.12	0.62

† Not applicable.

Rounds to zero.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-11. Standard errors for tables in appendix D: Percentage of students identified as an English language learner without a disability, by state: 2000 and 2003

State/jurisdiction	Grade 4		Grade 8	
	2000	2003	2000	2003
Alabama	0.12	0.14	0.25	0.26
Alaska	†	1.49	†	0.86
Arizona	1.27	1.55	1.44	1.31
Arkansas	0.56	0.76	0.31	0.60
California	2.28	2.00	1.30	1.45
Colorado	†	0.95	†	0.54
Connecticut	0.50	0.45	0.62	0.44
Delaware	†	0.25	†	0.19
District of Columbia	1.06	0.43	1.32	0.40
Florida	†	0.87	†	0.67
Georgia	0.47	0.77	0.33	0.32
Hawaii	0.65	0.64	0.88	0.45
Idaho	0.93	0.86	0.72	0.52
Illinois	1.59	0.79	0.91	0.50
Indiana	0.41	0.48	0.34	0.34
Iowa	0.75	0.53	#	0.38
Kansas	1.53	0.44	0.71	0.53
Kentucky	0.33	0.22	0.44	0.29
Louisiana	0.20	0.21	0.29	0.18
Maine	0.49	0.15	0.17	0.15
Maryland	0.55	0.60	0.56	0.45
Massachusetts	0.97	0.66	0.63	0.37
Michigan	0.64	1.10	0.26	0.78
Minnesota	0.96	0.39	1.45	0.41
Mississippi	#	0.10	0.05	0.23
Missouri	0.41	0.49	0.20	0.22
Montana	0.34	0.59	0.16	0.42
Nebraska	1.07	0.85	0.63	0.33
Nevada	1.09	1.26	0.88	0.52
New Hampshire	†	0.33	†	0.22
New Jersey	†	0.70	†	0.47
New Mexico	1.97	1.83	1.28	1.00
New York	0.90	0.94	0.98	0.41
North Carolina	0.45	0.55	0.48	0.52
North Dakota	0.61	0.69	0.21	0.43
Ohio	0.24	0.22	0.51	0.25
Oklahoma	0.91	1.13	0.62	0.64
Oregon	0.83	1.16	0.89	0.63
Pennsylvania	†	0.51	†	0.38
Rhode Island	0.95	0.87	0.78	0.40
South Carolina	0.17	0.32	0.20	0.18
South Dakota	†	0.82	†	0.71
Tennessee	0.24	0.35	0.51	0.46
Texas	1.17	1.42	1.02	0.49
Utah	0.77	1.20	0.53	0.65
Vermont	0.36	0.32	0.45	0.13
Virginia	0.83	1.21	0.46	0.28
Washington	†	0.67	†	0.58
West Virginia	#	0.07	0.08	0.11
Wisconsin	0.95	1.12	0.45	0.66
Wyoming	0.88	0.28	0.73	0.21

† Not applicable.

Rounds to zero.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-12. Standard errors for tables in appendix D: Percentage of students identified either with a disability or as an English language learner or both, by state: 2000 and 2003

State/jurisdiction	Grade 4		Grade 8	
	2000	2003	2000	2003
Alabama	0.82	0.62	0.85	0.91
Alaska	†	1.44	†	1.02
Arizona	1.36	1.42	1.56	1.36
Arkansas	0.93	1.14	0.91	1.08
California	2.42	2.05	1.63	1.49
Colorado	†	1.05	†	0.67
Connecticut	0.90	0.84	1.03	0.67
Delaware	†	0.59	†	0.77
District of Columbia	1.13	0.74	1.74	0.89
Florida	†	1.24	†	0.96
Georgia	0.69	0.97	0.66	0.69
Hawaii	0.98	0.89	1.86	0.72
Idaho	1.24	1.09	0.82	0.83
Illinois	1.57	0.98	1.22	0.86
Indiana	0.83	0.91	0.85	0.71
Iowa	1.07	0.99	#	0.85
Kansas	1.72	0.75	0.99	0.75
Kentucky	0.77	0.80	0.63	0.81
Louisiana	0.98	1.46	0.88	1.02
Maine	0.93	0.96	0.82	0.80
Maryland	0.89	1.05	0.82	1.01
Massachusetts	1.30	0.99	1.04	0.81
Michigan	0.86	1.18	0.76	0.88
Minnesota	1.38	0.76	1.23	0.62
Mississippi	0.53	0.56	0.73	0.55
Missouri	0.71	0.93	0.91	0.75
Montana	1.41	0.94	0.90	0.72
Nebraska	1.77	1.16	0.99	0.69
Nevada	1.17	1.19	0.86	0.89
New Hampshire	†	0.72	†	0.70
New Jersey	†	1.19	†	1.00
New Mexico	1.85	1.57	1.48	1.42
New York	1.37	1.18	1.16	0.72
North Carolina	0.94	0.96	0.83	0.80
North Dakota	0.91	1.01	0.84	0.88
Ohio	1.07	0.73	0.63	0.77
Oklahoma	1.28	1.32	0.83	1.12
Oregon	1.16	1.33	1.42	0.96
Pennsylvania	†	1.15	†	0.86
Rhode Island	1.19	1.28	1.38	0.75
South Carolina	0.81	0.82	0.68	0.86
South Dakota	†	1.06	†	0.97
Tennessee	0.97	0.93	0.84	0.81
Texas	1.41	1.43	1.02	0.94
Utah	0.91	1.18	1.03	0.93
Vermont	1.50	0.81	1.16	0.76
Virginia	1.27	1.50	0.86	0.81
Washington	†	1.08	†	0.86
West Virginia	0.78	0.89	0.77	0.89
Wisconsin	1.36	1.33	0.97	0.92
Wyoming	1.25	0.71	1.14	0.65

† Not applicable.

Rounds to zero.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-13. Standard errors for tables in appendix D: Percentage of students identified both with a disability and as an English language learner, and excluded, by state: 2000 and 2003

State/jurisdiction	Grade 4		Grade 8	
	2000	2003	2000	2003
Alabama	#	0.04	0.18	0.06
Alaska	†	0.08	†	0.04
Arizona	0.31	0.35	0.21	0.20
Arkansas	0.12	0.10	#	0.07
California	0.19	0.18	0.21	0.18
Colorado	†	0.06	†	0.09
Connecticut	0.13	0.12	0.17	0.14
Delaware	†	0.09	†	0.13
District of Columbia	#	0.13	0.20	0.14
Florida	†	0.16	†	0.14
Georgia	0.04	0.10	0.17	0.10
Hawaii	0.21	0.13	0.15	0.12
Idaho	0.29	0.08	0.04	0.09
Illinois	0.21	0.20	0.17	0.14
Indiana	0.14	0.11	#	0.11
Iowa	†	0.19	†	#
Kansas	0.21	0.08	#	0.14
Kentucky	0.09	0.11	0.16	0.08
Louisiana	0.13	0.04	#	0.13
Maine	#	0.22	0.08	0.11
Maryland	0.08	0.28	0.21	0.09
Massachusetts	#	0.10	0.20	0.15
Michigan	0.08	0.10	#	0.09
Minnesota	0.17	0.07	0.16	0.07
Mississippi	#	0.20	0.07	#
Missouri	0.07	0.05	#	0.09
Montana	#	0.09	0.21	0.07
Nebraska	#	0.19	0.18	0.06
Nevada	0.20	0.19	0.13	0.14
New Hampshire	†	0.07	†	0.03
New Jersey	†	0.19	†	0.09
New Mexico	0.51	0.28	0.56	0.19
New York	0.32	0.29	0.32	0.13
North Carolina	0.11	0.11	0.30	0.14
North Dakota	#	0.13	0.13	0.05
Ohio	#	0.21	0.59	0.06
Oklahoma	0.16	0.12	0.09	0.11
Oregon	0.24	0.18	0.13	0.14
Pennsylvania	†	0.11	†	0.07
Rhode Island	0.10	0.25	0.10	0.11
South Carolina	0.24	0.12	0.09	0.10
South Dakota	†	0.08	†	0.07
Tennessee	0.40	0.11	#	0.09
Texas	0.22	0.21	0.25	0.21
Utah	0.15	0.19	0.13	0.11
Vermont	0.08	0.10	0.11	0.09
Virginia	0.21	0.24	0.08	0.22
Washington	†	0.10	†	0.08
West Virginia	0.11	0.03	#	0.05
Wisconsin	0.17	0.11	0.09	0.10
Wyoming	#	0.05	#	0.11

† Not applicable.

Rounds to zero.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

National Assessment of Educational Progress

Table B-14. Standard errors for tables in appendix D: Percentage of students identified with a disability but not as an English language learner, and excluded, by state: 2000 and 2003

State/jurisdiction	Grade 4		Grade 8	
	2000	2003	2000	2003
Alabama	0.63	0.31	0.83	0.38
Alaska	†	0.25	†	0.18
Arizona	0.52	0.30	0.53	0.26
Arkansas	0.57	0.24	0.44	0.25
California	0.72	0.24	0.71	0.16
Colorado	†	0.25	†	0.22
Connecticut	0.89	0.49	0.63	0.34
Delaware	†	0.35	†	0.53
District of Columbia	0.42	0.30	1.29	0.37
Florida	†	0.27	†	0.22
Georgia	0.52	0.22	0.46	0.26
Hawaii	0.77	0.29	1.19	0.27
Idaho	0.09	0.19	0.33	0.11
Illinois	0.64	0.38	0.76	0.41
Indiana	0.56	0.38	0.64	0.27
Iowa	0.31	0.33	#	0.32
Kansas	0.49	0.25	0.56	0.29
Kentucky	0.68	0.46	0.71	0.61
Louisiana	0.47	0.88	0.55	0.59
Maine	0.64	0.42	0.49	0.39
Maryland	0.38	0.43	0.57	0.60
Massachusetts	0.22	0.31	0.40	0.43
Michigan	0.62	0.41	0.86	0.48
Minnesota	0.47	0.28	0.45	0.25
Mississippi	0.48	0.43	0.84	0.45
Missouri	0.50	0.29	0.59	0.56
Montana	0.67	0.28	0.54	0.26
Nebraska	0.51	0.31	1.23	0.34
Nevada	0.50	0.39	0.49	0.22
New Hampshire	†	0.43	†	0.49
New Jersey	†	0.47	†	0.24
New Mexico	0.65	0.28	0.66	0.24
New York	0.63	0.38	0.55	0.61
North Carolina	0.66	0.39	0.52	0.48
North Dakota	0.41	0.25	0.30	0.24
Ohio	1.14	0.55	0.86	0.81
Oklahoma	0.87	0.37	0.75	0.32
Oregon	0.44	0.36	0.35	0.29
Pennsylvania	†	0.45	†	0.29
Rhode Island	0.41	0.25	0.41	0.27
South Carolina	0.73	0.56	0.72	0.74
South Dakota	†	0.24	†	0.24
Tennessee	0.41	0.41	0.50	0.41
Texas	0.71	0.55	0.67	0.52
Utah	0.42	0.28	0.46	0.29
Vermont	0.61	0.31	0.56	0.35
Virginia	0.48	0.45	0.59	0.51
Washington	†	0.35	†	0.27
West Virginia	0.66	0.43	0.47	0.40
Wisconsin	1.06	0.41	0.88	0.39
Wyoming	0.79	0.13	0.36	0.15

† Not applicable.

Rounds to zero.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-15. Standard errors for tables in appendix D: Percentage of students identified as an English language learner without a disability, and excluded, by state: 2000 and 2003

State/jurisdiction	Grade 4		Grade 8	
	2000	2003	2000	2003
Alabama	#	0.04	0.07	0.16
Alaska	†	0.06	†	0.07
Arizona	0.77	0.22	0.45	0.29
Arkansas	0.07	0.33	0.29	0.20
California	0.65	0.36	0.60	0.32
Colorado	†	0.20	†	0.16
Connecticut	0.31	0.21	0.39	0.15
Delaware	†	0.16	†	0.14
District of Columbia	0.67	0.17	0.55	0.21
Florida	†	0.32	†	0.32
Georgia	0.35	0.16	0.30	0.12
Hawaii	0.72	0.39	0.59	0.17
Idaho	0.41	0.21	0.21	0.08
Illinois	0.52	0.43	0.98	0.18
Indiana	0.30	0.04	0.18	0.05
Iowa	0.35	0.22	#	0.10
Kansas	#	0.10	0.16	0.19
Kentucky	#	0.14	0.28	0.13
Louisiana	0.09	#	0.08	0.06
Maine	0.19	0.07	0.06	0.04
Maryland	0.37	0.28	0.44	0.33
Massachusetts	0.70	0.25	0.33	0.24
Michigan	0.62	0.20	0.26	0.21
Minnesota	0.30	0.12	0.34	0.18
Mississippi	#	0.09	0.05	0.10
Missouri	0.32	0.13	0.15	0.07
Montana	0.21	#	#	#
Nebraska	0.63	0.19	0.30	0.22
Nevada	0.85	0.38	0.32	0.11
New Hampshire	†	0.14	†	0.11
New Jersey	†	0.21	†	0.28
New Mexico	0.25	0.59	0.39	0.22
New York	0.75	0.35	0.42	0.25
North Carolina	0.21	0.13	0.32	0.15
North Dakota	0.09	0.04	#	#
Ohio	0.20	0.12	0.21	0.16
Oklahoma	0.24	0.18	0.17	0.09
Oregon	0.28	0.15	0.21	0.15
Pennsylvania	†	0.26	†	0.09
Rhode Island	0.20	0.45	0.36	0.20
South Carolina	0.09	0.08	0.15	0.07
South Dakota	†	0.05	†	0.03
Tennessee	0.09	0.14	0.27	0.23
Texas	0.47	0.17	0.28	0.22
Utah	0.20	0.20	0.14	0.12
Vermont	#	0.06	0.33	0.03
Virginia	0.35	0.44	0.31	0.14
Washington	†	0.19	†	0.13
West Virginia	#	0.03	0.07	0.05
Wisconsin	0.36	0.22	0.19	0.12
Wyoming	#	0.03	#	0.04

† Not applicable.

Rounds to zero.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

National Assessment of Educational Progress

Table B-16. Standard errors for tables in appendix D: Percentage of students identified with either a disability or as an English language learner, and excluded, by state: 2000 and 2003

State/jurisdiction	Grade 4		Grade 8	
	2000	2003	2000	2003
Alabama	0.63	0.31	0.79	0.42
Alaska	†	0.26	†	0.20
Arizona	1.31	0.53	0.80	0.55
Arkansas	0.56	0.43	0.55	0.37
California	1.08	0.54	0.92	0.40
Colorado	†	0.31	†	0.25
Connecticut	1.00	0.57	0.89	0.38
Delaware	†	0.38	†	0.55
District of Columbia	0.75	0.34	1.46	0.45
Florida	†	0.50	†	0.47
Georgia	0.64	0.26	0.46	0.30
Hawaii	1.14	0.59	1.37	0.37
Idaho	0.48	0.29	0.34	0.17
Illinois	0.85	0.64	1.38	0.53
Indiana	0.69	0.41	0.69	0.35
Iowa	0.52	0.47	#	0.33
Kansas	0.52	0.26	0.61	0.41
Kentucky	0.69	0.50	0.74	0.70
Louisiana	0.52	0.88	0.57	0.62
Maine	0.64	0.51	0.50	0.41
Maryland	0.64	0.67	0.77	0.73
Massachusetts	0.78	0.41	0.57	0.58
Michigan	0.85	0.44	0.94	0.59
Minnesota	0.61	0.31	0.55	0.31
Mississippi	0.48	0.45	0.83	0.47
Missouri	0.66	0.34	0.60	0.59
Montana	0.70	0.28	0.57	0.26
Nebraska	0.63	0.43	1.19	0.42
Nevada	1.01	0.56	0.63	0.23
New Hampshire	†	0.44	†	0.50
New Jersey	†	0.60	†	0.42
New Mexico	1.00	0.85	0.89	0.41
New York	1.16	0.62	0.83	0.74
North Carolina	0.67	0.38	0.62	0.48
North Dakota	0.43	0.28	0.31	0.23
Ohio	1.11	0.53	0.93	0.82
Oklahoma	0.82	0.45	0.81	0.35
Oregon	0.65	0.48	0.42	0.39
Pennsylvania	†	0.63	†	0.35
Rhode Island	0.52	0.58	0.61	0.30
South Carolina	0.67	0.59	0.73	0.76
South Dakota	†	0.26	†	0.25
Tennessee	0.51	0.50	0.57	0.41
Texas	0.90	0.63	0.64	0.61
Utah	0.50	0.43	0.53	0.36
Vermont	0.59	0.38	0.70	0.37
Virginia	0.69	0.77	0.60	0.56
Washington	†	0.47	†	0.30
West Virginia	0.65	0.44	0.48	0.41
Wisconsin	1.26	0.45	0.98	0.44
Wyoming	0.79	0.13	0.36	0.20

† Not applicable.

Rounds to zero.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-17. Standard errors for tables in appendix D: Percentage of students identified both with a disability and as an English language learner, and accommodated, by state: 2000 and 2003

State/jurisdiction	Grade 4		Grade 8	
	2000	2003	2000	2003
Alabama	#	0.07	0.07	0.11
Alaska	†	0.49	†	0.22
Arizona	0.26	0.19	0.24	0.21
Arkansas	#	0.04	#	0.10
California	0.13	0.16	0.33	0.11
Colorado	†	0.23	†	0.12
Connecticut	#	0.10	#	0.11
Delaware	†	0.13	†	0.11
District of Columbia	0.07	0.14	0.20	0.16
Florida	†	0.33	†	0.20
Georgia	#	0.14	#	0.10
Hawaii	0.13	0.14	#	0.12
Idaho	0.17	0.21	0.34	0.11
Illinois	0.24	0.19	#	0.18
Indiana	0.07	0.13	#	0.23
Iowa	#	0.16	#	0.09
Kansas	#	0.10	#	0.14
Kentucky	#	0.13	#	0.08
Louisiana	0.13	0.26	#	0.15
Maine	#	0.06	#	0.09
Maryland	#	0.12	0.16	0.07
Massachusetts	0.24	0.15	0.11	0.10
Michigan	#	0.24	#	#
Minnesota	0.14	0.13	#	0.08
Mississippi	#	#	#	#
Missouri	#	0.16	#	0.07
Montana	#	0.30	#	0.14
Nebraska	#	0.16	#	0.08
Nevada	0.15	0.27	0.07	0.16
New Hampshire	†	0.12	†	0.11
New Jersey	†	0.18	†	0.17
New Mexico	0.54	0.36	0.25	0.42
New York	0.21	0.19	0.07	0.15
North Carolina	#	0.20	#	0.18
North Dakota	0.10	0.15	#	0.29
Ohio	#	0.17	#	0.04
Oklahoma	0.30	0.21	#	0.20
Oregon	0.10	0.21	0.13	0.16
Pennsylvania	†	0.15	†	0.11
Rhode Island	0.19	0.28	0.06	0.15
South Carolina	0.11	0.04	#	0.08
South Dakota	†	0.12	†	0.14
Tennessee	#	0.08	#	#
Texas	0.15	0.12	0.09	0.06
Utah	0.18	0.20	0.07	0.18
Vermont	#	0.12	#	0.09
Virginia	#	0.21	0.11	0.16
Washington	†	0.31	†	0.18
West Virginia	0.07	0.10	0.11	0.04
Wisconsin	0.23	0.18	0.13	0.15
Wyoming	#	0.18	#	0.16

† Not applicable.

Rounds to zero.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-18. Standard errors for tables in appendix D: Percentage of students identified with a disability but not as an English language learner, and accommodated, by state: 2000 and 2003

State/jurisdiction	Grade 4		Grade 8	
	2000	2003	2000	2003
Alabama	0.79	0.34	0.29	0.52
Alaska	†	0.60	†	0.49
Arizona	1.06	0.42	0.68	0.47
Arkansas	0.80	0.61	0.78	0.69
California	0.38	0.19	0.62	0.25
Colorado	†	0.57	†	0.53
Connecticut	0.69	0.56	0.60	0.57
Delaware	†	0.38	†	0.45
District of Columbia	0.80	0.43	0.85	0.54
Florida	†	0.71	†	0.74
Georgia	0.53	0.56	0.57	0.49
Hawaii	0.40	0.45	0.47	0.46
Idaho	0.92	0.47	0.52	0.38
Illinois	0.91	0.64	0.94	0.72
Indiana	1.18	0.55	0.70	0.67
Iowa	1.01	0.74	#	0.66
Kansas	0.69	0.64	0.95	0.58
Kentucky	0.90	0.52	0.68	0.49
Louisiana	1.07	1.25	1.01	0.73
Maine	0.90	0.69	0.75	0.69
Maryland	0.75	0.57	0.80	0.71
Massachusetts	0.99	0.86	1.06	0.67
Michigan	1.02	0.48	0.99	0.62
Minnesota	0.86	0.45	0.65	0.65
Mississippi	0.48	0.29	0.40	0.27
Missouri	0.92	0.63	1.11	0.62
Montana	1.21	0.61	0.98	0.43
Nebraska	1.26	0.66	0.82	0.48
Nevada	0.56	0.33	0.82	0.43
New Hampshire	†	0.75	†	0.52
New Jersey	†	0.64	†	0.77
New Mexico	0.99	0.57	0.52	0.52
New York	1.02	0.71	0.93	0.66
North Carolina	0.88	0.67	1.02	0.67
North Dakota	0.56	0.58	0.93	0.44
Ohio	1.06	0.58	0.78	0.72
Oklahoma	0.63	0.77	0.88	0.65
Oregon	1.06	0.60	1.06	0.53
Pennsylvania	†	0.63	†	0.64
Rhode Island	0.99	0.85	0.99	0.58
South Carolina	0.78	0.43	0.75	0.56
South Dakota	†	0.46	†	0.42
Tennessee	0.38	0.54	0.29	0.28
Texas	0.61	0.43	0.29	0.41
Utah	0.42	0.41	0.52	0.40
Vermont	1.26	0.69	1.01	0.62
Virginia	0.63	0.50	0.60	0.50
Washington	†	0.53	†	0.51
West Virginia	1.06	0.78	1.06	0.76
Wisconsin	1.14	0.77	0.94	0.66
Wyoming	1.14	0.51	0.84	0.46

† Not applicable.

Rounds to zero.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-19. Standard errors for tables in appendix D: Percentage of students identified as an English language learner without a disability, and accommodated, by state: 2000 and 2003

State/jurisdiction	Grade 4		Grade 8	
	2000	2003	2000	2003
Alabama	#	0.03	#	0.04
Alaska	†	0.29	†	0.10
Arizona	1.33	0.40	1.04	0.37
Arkansas	0.17	0.15	#	0.25
California	2.22	0.58	0.96	0.13
Colorado	†	0.76	†	0.36
Connecticut	0.36	0.42	0.38	0.45
Delaware	†	0.15	†	0.07
District of Columbia	0.87	0.28	0.75	0.24
Florida	†	0.38	†	0.39
Georgia	0.10	0.25	#	0.11
Hawaii	0.11	0.36	0.14	0.21
Idaho	0.34	0.14	0.26	0.20
Illinois	1.03	0.43	0.30	0.42
Indiana	0.31	0.26	#	0.19
Iowa	0.23	0.27	#	0.16
Kansas	0.41	0.36	#	0.44
Kentucky	#	0.03	0.07	0.11
Louisiana	0.11	0.23	0.29	0.09
Maine	#	0.05	0.16	0.07
Maryland	0.06	0.17	0.15	0.09
Massachusetts	1.00	0.24	0.27	0.15
Michigan	#	0.23	#	0.70
Minnesota	0.97	0.34	0.21	0.20
Mississippi	#	#	#	#
Missouri	0.12	0.35	#	0.19
Montana	#	0.27	#	0.26
Nebraska	0.29	0.20	0.16	0.13
Nevada	0.34	0.51	0.22	0.20
New Hampshire	†	0.10	†	0.16
New Jersey	†	0.60	†	0.34
New Mexico	1.60	0.86	0.61	0.64
New York	0.52	0.49	0.50	0.34
North Carolina	0.31	0.32	0.13	0.28
North Dakota	0.08	0.06	#	0.12
Ohio	#	0.10	0.10	0.07
Oklahoma	0.57	0.20	0.06	0.30
Oregon	0.86	0.72	0.32	0.53
Pennsylvania	†	0.22	†	0.20
Rhode Island	0.55	0.58	0.30	0.24
South Carolina	#	0.11	0.08	0.07
South Dakota	†	0.62	†	0.59
Tennessee	#	0.03	0.10	0.08
Texas	1.00	0.60	0.33	0.15
Utah	0.52	0.55	0.24	0.25
Vermont	0.36	0.16	0.19	#
Virginia	0.46	0.64	0.19	0.20
Washington	†	0.29	†	0.16
West Virginia	#	#	#	#
Wisconsin	0.81	0.60	0.23	0.34
Wyoming	0.11	0.08	#	0.09

† Not applicable.

Rounds to zero.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table B-20. Standard errors for tables in appendix D: Percentage of students identified with either a disability or as an English language learner or both, and accommodated, by state: 2000 and 2003

State/jurisdiction	Grade 4		Grade 8	
	2000	2003	2000	2003
Alabama	0.79	0.36	0.30	0.54
Alaska	†	0.65	†	0.57
Arizona	1.88	0.65	1.30	0.73
Arkansas	0.81	0.64	0.78	0.71
California	2.33	0.69	1.06	0.33
Colorado	†	1.11	†	0.54
Connecticut	0.80	0.71	0.64	0.67
Delaware	†	0.39	†	0.46
District of Columbia	1.19	0.51	1.27	0.63
Florida	†	0.86	†	0.74
Georgia	0.53	0.67	0.57	0.54
Hawaii	0.48	0.57	0.57	0.51
Idaho	0.94	0.55	0.91	0.50
Illinois	1.37	0.82	0.91	0.86
Indiana	1.25	0.65	0.70	0.74
Iowa	1.03	0.76	#	0.67
Kansas	0.73	0.71	0.95	0.73
Kentucky	0.90	0.54	0.66	0.46
Louisiana	1.06	1.24	1.05	0.75
Maine	0.90	0.68	0.77	0.66
Maryland	0.77	0.63	0.85	0.73
Massachusetts	1.34	0.87	1.08	0.66
Michigan	1.02	0.66	0.99	0.86
Minnesota	1.30	0.62	0.69	0.72
Mississippi	0.48	0.29	0.40	0.27
Missouri	0.94	0.74	1.11	0.62
Montana	1.21	0.74	0.98	0.52
Nebraska	1.38	0.69	0.84	0.51
Nevada	0.66	0.71	0.83	0.49
New Hampshire	†	0.80	†	0.56
New Jersey	†	0.89	†	0.89
New Mexico	2.15	1.09	0.77	0.99
New York	1.21	0.88	1.16	0.66
North Carolina	0.95	0.77	1.01	0.65
North Dakota	0.52	0.60	0.93	0.52
Ohio	1.06	0.61	0.82	0.73
Oklahoma	0.84	0.85	0.88	0.80
Oregon	1.44	0.97	1.21	0.86
Pennsylvania	†	0.72	†	0.66
Rhode Island	1.15	1.15	1.11	0.65
South Carolina	0.81	0.44	0.71	0.56
South Dakota	†	0.77	†	0.71
Tennessee	0.38	0.52	0.30	0.29
Texas	1.22	0.77	0.44	0.45
Utah	0.69	0.66	0.53	0.53
Vermont	1.25	0.67	1.14	0.61
Virginia	0.93	0.87	0.66	0.55
Washington	†	0.73	†	0.62
West Virginia	1.06	0.77	1.04	0.74
Wisconsin	1.54	0.92	1.03	0.88
Wyoming	1.18	0.56	0.84	0.50

† Not applicable.

Rounds to zero.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

C Standard NAEP Estimates

The tables in this appendix are based on standard NAEP estimates, which do not include representation of the mathematics achievement of those students with disabilities and English language learners who are excluded from NAEP testing sessions or would be excluded if selected. All other achievement tables and figures in this report are based on full population estimates, which include representation of the achievement of excluded students. The method for estimation of excluded students' mathematics achievement is described in appendix A.

Table C-1. NAEP equivalent of state grade 4 mathematics achievement standards, by state: 2003 (corresponds to table B-3)

State/ jurisdiction	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5
Alabama	—	—	—	—	—
Alaska	—	—	222.9	—	—
Arizona	—	194.2	231.5	240.7	—
Arkansas	—	209.4	223.6	238.9	—
California	180.2	209.7	231.1	255.3	—
Colorado	—	—	201.8	232.4	260.9
Connecticut	204.3	217.1	236.1	263.5	—
Delaware	—	202.3	219.9	258.5	272.0
District of Columbia	—	192.4	218.6	250.1	—
Florida	—	211.5	230.7	258.1	281.1
Georgia	—	—	211.9	256.7	—
Hawaii	—	206.4	251.7	283.1	—
Idaho	—	185.3	216.9	254.1	—
Illinois	—	176.2	217.8	272.3	—
Indiana	—	—	225.7	270.4	—
Iowa	—	—	221.8	—	—
Kansas	—	203.5	225.9	241.9	263.4
Kentucky	—	217.4	237.6	265.2	—
Louisiana	202.4	221.3	255.1	276.7	—
Maine	—	223.4	252.3	281.4	—
Maryland	—	—	228.7	274.2	—
Massachusetts	—	214.9	251.2	274.3	—
Michigan	—	192.3	226.5	259.4	—
Minnesota	194.9	220.9	237.5	267.7	—
Mississippi	—	188.6	205.4	—	—
Missouri	181.8	212.8	244.4	273.2	—
Montana	—	205.4	219.9	260.9	—
Nebraska	—	—	—	—	—
Nevada	—	207.9	228.2	248.6	—
New Hampshire	—	—	221.7	248.7	269.5
New Jersey	—	—	227.1	259.6	—
New Mexico	—	206.9	229.0	247.1	—
New York	—	187.2	213.5	250.5	—
North Carolina	—	187.9	203.5	236.7	—
North Dakota	—	—	233.7	—	—
Ohio	—	222.7	232.4	265.0	—
Oklahoma	—	174.3	217.2	256.4	—
Oregon	—	—	215.5	255.0	—
Pennsylvania	—	213.2	233.2	254.6	—
Rhode Island	—	—	237.8	—	—
South Carolina	—	213.9	248.5	266.2	—
South Dakota	—	171.3	223.4	258.6	—
Tennessee	—	—	—	—	—
Texas	—	—	207.0	—	—
Utah	—	—	—	—	—
Vermont	185.4	226.6	241.3	268.9	—
Virginia	—	—	221.9	264.8	—
Washington	—	215.1	236.0	257.1	—
West Virginia	—	—	—	—	—
Wisconsin	—	212.2	223.1	253.2	—
Wyoming	—	228.9	250.2	273.1	—

— Not available.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table C-2. NAEP equivalent of state grade 8 mathematics achievement standards, by state: 2003 (corresponds to table B-3)

State/jurisdiction	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5
Alabama	—	—	—	—	—
Alaska	—	—	268.0	—	—
Arizona	—	263.2	299.9	321.5	—
Arkansas	—	251.6	295.8	332.2	—
California	216.3	253.0	287.1	323.0	—
Colorado	—	—	267.7	294.4	322.4
Connecticut	235.5	258.0	279.1	313.0	—
Delaware	—	255.1	278.2	306.5	320.1
District of Columbia	—	251.6	282.9	322.3	—
Florida	—	245.6	268.9	299.6	319.9
Georgia	—	—	255.2	307.7	—
Hawaii	—	246.3	299.3	345.5	—
Idaho	—	247.6	280.1	315.8	—
Illinois	—	217.6	275.7	313.7	—
Indiana	—	—	269.4	322.3	—
Iowa	—	—	266.6	—	—
Kansas	—	255.9	278.3	296.4	317.1
Kentucky	—	253.9	291.2	320.8	—
Louisiana	245.1	265.4	310.7	325.9	—
Maine	—	269.1	311.2	358.5	—
Maryland	—	—	286.0	318.6	—
Massachusetts	—	270.6	298.9	326.2	—
Michigan	—	254.1	278.1	298.2	—
Minnesota	—	—	—	—	—
Mississippi	—	241.6	261.4	—	—
Missouri	247.2	281.4	314.4	347.6	—
Montana	—	254.5	271.0	318.4	—
Nebraska	—	—	—	—	—
Nevada	—	251.8	277.2	300.0	—
New Hampshire	—	—	—	—	—
New Jersey	—	—	278.5	317.6	—
New Mexico	—	242.2	269.6	294.5	—
New York	—	242.1	279.2	324.5	—
North Carolina	—	220.8	246.8	284.8	—
North Dakota	—	—	293.4	—	—
Ohio	—	—	—	—	—
Oklahoma	—	215.4	256.0	304.4	—
Oregon	—	—	275.3	298.9	—
Pennsylvania	—	256.4	279.4	309.8	—
Rhode Island	—	—	287.0	—	—
South Carolina	—	263.1	306.0	327.4	—
South Dakota	—	225.1	281.6	329.0	—
Tennessee	—	—	—	—	—
Texas	—	—	259.0	—	—
Utah	—	—	—	—	—
Vermont	242.4	271.0	286.0	316.9	—
Virginia	—	—	255.8	313.7	—
Washington	—	277.9	295.0	311.6	—
West Virginia	—	—	—	—	—
Wisconsin	—	242.3	260.9	305.9	—
Wyoming	—	267.8	296.9	321.0	—

— Not available.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table C-3. Correlations between NAEP and state assessment school-level percentages meeting primary state mathematics standards, grades 4 and 8, by state: 2003 (corresponds to table 4)

State/ jurisdiction	Name of standard	Grades		Grade 4 standard error	Grade 8 standard error
Alabama	Percentile Rank	4	8	0.80	0.84
Alaska	Proficient	4	8	0.79	0.86
Arizona	Meeting	5	8	0.77	0.69
Arkansas	Proficient	4	8	0.82	0.77
California	Proficient	4	7	0.84	0.88
Colorado	Partially Proficient	5	8	0.80	0.87
Connecticut	Goal	4	8	0.88	0.89
Delaware	Meeting	5	8	0.62	0.80
District of Columbia	Proficient	4	8	0.68	0.97
Florida	(3) Partial Success	4	8	0.88	0.85
Georgia	Meeting	4	8	0.83	0.81
Hawaii	Meeting	5	8	0.78	0.82
Idaho	Proficient	4	8	0.67	0.70
Illinois	Meeting	5	8	0.84	0.91
Indiana	Pass	3	8	0.44	0.82
Iowa	Proficient	4	8	0.78	0.77
Kansas	Proficient	4	7	0.65	0.72
Kentucky	Proficient	5	8	0.53	0.70
Louisiana	Mastery	4	8	0.79	0.81
Maine	Meeting	4	8	0.56	0.70
Maryland	Proficient	5	8	0.82	0.87
Massachusetts	Proficient	4	8	0.82	0.87
Michigan	Meeting	4	8	0.75	0.86
Minnesota	(3) Proficient	5	—	0.75	—
Mississippi	Proficient	4	8	0.81	0.83
Missouri	Proficient	4	8	0.69	0.62
Montana	Proficient	4	8	0.73	0.71
Nebraska	Meeting	—	—	—	—
Nevada	Meeting:3	4	7	0.81	0.81
New Hampshire	Basic	3	—	0.43	—
New Jersey	Proficient	4	8	0.84	0.90
New Mexico	Top half	4	8	0.77	0.81
New York	Meeting	4	8	0.86	0.86
North Carolina	Consistent Mastery	4	8	0.57	0.70
North Dakota	Meeting	4	8	0.64	0.76
Ohio	Proficient	4	—	0.81	—
Oklahoma	Satisfactory	5	8	0.56	0.71
Oregon	Meeting	5	8	0.50	0.75
Pennsylvania	Proficient	5	8	0.84	0.86
Rhode Island	Proficient	4	8	0.76	0.90
South Carolina	Proficient	4	8	0.73	0.79
South Dakota	Proficient	4	8	0.77	0.71
Tennessee	Percentile Rank	4	8	0.75	0.80
Texas	Passing	4	8	0.56	0.70
Utah	Percentile Rank	5	8	0.66	0.70
Vermont	Achieved	4	8	0.48	0.75
Virginia	Proficient	5	8	0.55	0.65
Washington	Met	4	7	0.68	0.67
West Virginia	Top half	—	—	—	—
Wisconsin	Proficient	4	8	0.82	0.89
Wyoming	Proficient	4	8	0.64	0.74

— Not available.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table C-4. Mathematics achievement gains in percentage meeting the primary standard in grade 4 (corresponds to table 8)

State/ jurisdiction	State			NAEP		
	2000	2003	Gain	2000	2003	Gain
Alabama	57.5	53.1	-4.4	41.2	35.7	-5.5
Alaska	—	67.3	—	—	67.3	—
Arizona	40.9	49.8	8.9*	41.0	55.5	14.5*
Arkansas	35.8	60.1	24.3*	35.8	52.3	16.5*
California	52.4	—	—	52.2	—	—
Colorado	—	86.2	—	—	86.2	—
Connecticut	63.9	58.9	-5.0*	63.9	71.7	7.8*
Delaware	—	72.6	—	—	72.5	—
District of Columbia	37.0	—	—	36.9	—	—
Florida	—	55.6	—	—	55.5	—
Georgia	61.6	73.9	12.3	61.6	76.8	15.2
Hawaii	65.2	67.6	2.4*	65.2	77.1	11.9*
Idaho	—	76.7	—	—	76.7	—
Illinois	53.2	68.6	15.4	53.2	66.0	12.8
Indiana	—	—	—	—	—	—
Iowa	—	75.2	—	—	75.2	—
Kansas	59.2	73.9	14.7	59.3	70.8	11.5
Kentucky	31.3	38.0	6.7	31.3	41.2	9.9
Louisiana	11.5	15.7	4.2*	11.6	19.8	8.2*
Maine	23.3	29.1	5.8*	23.3	34.4	11.1*
Maryland	45.6	—	—	45.6	—	—
Massachusetts	41.4	38.3	-3.1*	41.2	51.4	10.2*
Michigan	76.9	—	—	76.8	—	—
Minnesota	47.8	58.3	10.5*	47.8	57.7	9.9*
Mississippi	—	74.0	—	—	74.1	—
Missouri	36.5	36.8	0.3*	36.5	46.1	9.6*
Montana	—	75.3	—	—	75.3	—
Nebraska	60.0	—	—	59.8	—	—
Nevada	—	—	—	—	—	—
New Hampshire	—	80.4	—	—	80.3	—
New Jersey	—	67.9	—	—	67.9	—
New Mexico	—	42.3	—	—	42.3	—
New York	68.2	79.4	11.2	68.2	76.8	8.6
North Carolina	84.5	92.3	7.8	84.5	92.8	8.3
North Dakota	—	59.1	—	—	59.1	—
Ohio	42.6	59.3	16.7*	42.8	53.5	10.7*
Oklahoma	86.2	69.4	-16.8*	86.2	89.4	3.2*
Oregon	67.3	78.0	10.7	67.4	81.1	13.7
Pennsylvania	—	56.9	—	—	56.8	—
Rhode Island	20.8	42.5	21.7*	21.0	26.2	5.2*
South Carolina	22.9	32.8	9.9*	22.8	38.9	16.1*
South Dakota	—	72.5	—	—	72.6	—
Tennessee	—	54.5	—	—	41.3	—
Texas	88.7	—	—	88.7	—	—
Utah	50.2	48.1	-2.1	52.3	50.8	-1.5
Vermont	47.0	53.0	6.0*	47.0	62.2	15.2*
Virginia	—	—	—	—	—	—
Washington	—	54.1	—	—	54.1	—
West Virginia	—	—	—	—	—	—
Wisconsin	72.2	—	—	72.3	—	—
Wyoming	25.6	36.4	10.8*	25.6	41.2	15.6*

— Not available.

* State and NAEP gains are significantly different from each other (*p*<.05).

NOTE: Primary standard is the state's standard for *proficient* performance.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table C-5. Mathematics achievement gains in percentage meeting the primary standard in grade 8: 2000 and 2003 (corresponds to table 9)

State/ jurisdiction	State			NAEP		
	2000	2003	Gain	2000	2003	Gain
Alabama	55.7	50.7	-5.0	41.9	38.3	-3.6
Alaska	—	65.2	—	—	65.4	—
Arizona	18.2	20.8	2.6	18.1	18.1	0.0
Arkansas	13.3	21.9	8.6*	13.3	18.3	5.0*
California	48.1	—	—	48.1	—	—
Colorado	—	68.4	—	—	68.4	—
Connecticut	58.3	56.3	-2.0*	58.4	60.1	1.7*
Delaware	—	48.2	—	—	48.3	—
District of Columbia	8.5	—	—	8.4	—	—
Florida	—	54.3	—		54.3	—
Georgia	54.9	66.4	11.5*	54.9	59.7	4.8*
Hawaii	60.7	54.2	-6.5*	60.7	64.9	4.2*
Idaho	—	52.4	—	—	52.4	
Illinois	46.1	54.0	7.9*	46.1	49.3	3.2*
Indiana	—	—	—	—	—	—
Iowa	—	71.7	—	—	71.8	—
Kansas	55.2	59.4	4.2	55.3	55.6	0.3
Kentucky	26.2	32.0	5.8	26.1	30.3	4.2
Louisiana	7.3	8.8	1.5	7.4	10.9	3.5
Maine	20.6	17.1	-3.5	20.8	19.3	-1.5
Maryland	51.3	—	—	51.3	—	—
Massachusetts	34.0	38.4	4.4*	34.1	41.9	7.8*
Michigan	66.9	—	—	66.9	—	—
Minnesota	—	—	—	—	—	—
Mississippi	—	46.1	—	—	46.2	
Missouri	12.9	13.4	0.5*	12.9	18.1	5.2*
Montana	—	70.4	—	—	70.4	
Nebraska	60.4	—	—	61.5	—	—
Nevada	—	—	—	—	—	—
New Hampshire	—	—	—	—	—	—
New Jersey	—	56.4	—	—	56.4	—
New Mexico	—	39.5	—	—	39.6	—
New York	41.3	54.1	12.8*	41.4	48.9	7.5*
North Carolina	80.8	82.4	1.6	80.9	83.5	2.6
North Dakota	—	43.6	—	—	43.7	—
Ohio	—	—	—	—	—	—
Oklahoma	70.8	71.3	0.5	70.9	73.3	2.4
Oregon	54.8	57.7	2.9	54.9	55.1	0.2
Pennsylvania	—	51.6	—	—	51.6	
Rhode Island	26.8	35.6	8.8*	26.9	29.1	2.2*
South Carolina	19.4	20.4	1.0*	19.3	29.3	10.*
South Dakota	—	57.5	—	—	57.6	—
Tennessee	—	57.8	—	—	41.9	—
Texas	89.9	—	—	89.9		—
Utah	57.9	56.9	-1.0	52.0	53.5	1.5
Vermont	45.8	51.6	5.8	45.8	52.4	6.6
Virginia	—	—	—	—	—	—
Washington	—	36.6	—	—	36.7	—
West Virginia	—	—	—	—	—	—
Wisconsin	43.1	—	—	43.3	—	—
Wyoming	31.5	35.1	3.6*	31.5	41.3	9.8*

— Not available.

* State and NAEP gains are significantly different from each other (*p*<.05).

NOTE: Primary standard is the state's standard for *proficient* performance.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2003 Mathematics Assessments: The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table C-6. Differences between NAEP and state assessments of grade 4 mathematics achievement race and poverty gaps, by state: 2003 (corresponds to table 10)

State/ jurisdiction	Black-White	Hispanic-White	Poverty
Alabama	-7.4 *	—	-4.2 *
Alaska	—	—	—
Arizona	—	-12.6 *	—
Arkansas	-8.7 *	—	1.2
California	—	-11.5 *	-7.5 *
Colorado	—	—	—
Connecticut	-9.7 *	-5.8	-6.2 *
Delaware	-4.4 *	—	-0.4
District of Columbia	—	—	-16.5 *
Florida	-8.7 *	-2.6	-3.1
Georgia	-8.8 *	—	-3.5
Hawaii	—	—	-5.4 *
Idaho	—	-10.6 *	—
Illinois	-3.0	-6.6 *	-3.6
Indiana	-23.6 *	—	-11.8 *
Iowa	—	—	—
Kansas	-11.8 *	—	-7.9
Kentucky	-2.1	—	-2.0
Louisiana	-4.5 *	—	-3.1
Maine	—	—	—
Maryland	—	—	—
Massachusetts	-5.8 *	-6.7	—
Michigan	—	—	—
Minnesota	—	—	-2.4
Mississippi	-2.7	—	-3.2
Missouri	-13.5 *	—	-4.0
Montana	—	—	—
Nebraska	—	—	—
Nevada	—	-6.1 *	-6.7 *
New Hampshire	—	—	-8.8 *
New Jersey	-8.3	-14.0 *	-5.0
New Mexico	—	-8.0 *	-2.1
New York	-5.7 *	-4.9	-2.0
North Carolina	-5.8 *	—	-2.2
North Dakota	—	—	—
Ohio	-9.9 *	—	-7.7 *
Oklahoma	-6.7	—	—
Oregon	—	—	—
Pennsylvania	-5.5	—	-3.5
Rhode Island	—	-13.4 *	
South Carolina	-5.7 *	—	-2.4
South Dakota	—	—	-3.2
Tennessee	-3.7	—	-0.5
Texas	-3.8	-2.6	—
Utah	—	—	—
Vermont	—	—	-13.1 *
Virginia	-12.5 *	—	—
Washington	—	0.5	—
West Virginia	—	—	—
Wisconsin	-9.6*	—	-8.9 *
Wyoming	—	—	-4.6 *

— Not available.

* State and NAEP gains are significantly different from each other (*p*<.05).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Table C-7. Differences between NAEP and state assessments of grade 8 mathematics achievement race and poverty gaps, by state: 2003 (corresponds to table 11)

State/jurisdiction	Black-White	Hispanic-White	Poverty
Alabama	-8.0 *	—	-3.5
Alaska	—	—	—
Arizona	—	-5.2 *	—
Arkansas	1.0		4.2
California	—	-4.1	-2.0
Colorado	—	—	—
Connecticut	0.6	1.1	1.1
Delaware	-1.7	—	4.0 *
District of Columbia	—	—	-1.1
Florida	-7.5 *	-6.2 *	-1.1
Georgia	-13.5 *	—	-7.8 *
Hawaii	—	—	-4.3 *
Idaho	—	-6.0	
Illinois	-1.7	-5.3	-5.3 *
Indiana	-0.9	—	4.9
Iowa	—	—	—
Kansas	—	—	2.1
Kentucky	—	—	0.4
Louisiana	-1.0	—	-0.5
Maine	—	—	—
Maryland	—	—	—
Massachusetts		-4.6	
Michigan	—	—	—
Minnesota	—	—	—
Mississippi	-7.1 *	—	-3.0
Missouri	-0.8	—	-0.0
Montana	—	—	—
Nebraska	—	—	—
Nevada	—	-5.0 *	0.2
New Hampshire	—	—	—
New Jersey	-1.3	-1.5	-1.3
New Mexico	—	-7.2 *	-0.6
New York	-10.6 *	-5.4	-1.5
North Carolina	-5.2 *		-0.5
North Dakota	—	—	—
Ohio	—	—	—
Oklahoma	-0.8	—	—
Oregon		—	
Pennsylvania	-4.7	—	-3.5
Rhode Island		-3.1 *	
South Carolina	4.1 *	—	5.7 *
South Dakota	—	—	0.0
Tennessee	-2.9	—	3.6
Texas	-8.3 *	-3.6	—
Utah	—	—	—
Vermont	—	—	-3.3
Virginia	-10.2 *	—	—
Washington	—	—	—
West Virginia	—	—	—
Wisconsin	—	—	-2.1
Wyoming	—	—	5.3 *

— Not available.

* State and NAEP gains are significantly different from each other (*p*<.05).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.