NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

U.S. Department of Education
NCES 2008-475

Comparison Between NAEP and State Mathematics Assessment Results: 2003

Volume 1

Research and Development Report

U.S. Department of Education
NCES 2008-475

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

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Mark Schneider
Commissioner

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

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

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

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

January 2008

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

Suggested Citation
McLaughlin, D.H., 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 Mathematics Assessment Results: 2003* (NCES 2008-475). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

For ordering information on this report, write to
U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794-1398

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

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

Foreword

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

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

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

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

Executive Summary

In 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 particularly 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

Page

List of Tables

Table	Page

Table	Page

Table	Page

List of Figures

Figure	Page

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

	Grade 4			Grade 8		
State/ jurisdiction	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.