CENTER ON EDUCATION POLICY CEP

Has Student Achievement Increased Since 2002?

State Test Score Trends Through 2006–07

Center on Education Policy

JUNE 2008

CENTER ON EDUCATION POLICY CEP

Has Student Achievement Increased Since 2002?

State Test Score Trends Through 2006–07

Center on Education Policy

JUNE 2008

Table of Contents

Tables and Figures

Chapter 4

Chapter 5

Chapter 6

Appendix 1.

Appendix 2.

CHAPTER 1

Executive Summary

This report describes findings from the second year of the most comprehensive, intensive, and carefully constructed study to date of trends in student achievement in all 50 states since 2002, the year the No Child Left Behind Act (NCLB) was enacted.

Last year, the Center on Education Policy (CEP), an independent nonprofit organization, began this work by seeking to answer the question of whether student achievement has increased *because* of NCLB. We soon discovered that it is not possible to directly relate changes in student achievement to NCLB, as we explain later. However, it is possible to learn much more about student achievement now than it was before 2002, when NCLB was enacted, because the law has greatly expanded student testing, accountability, and reporting of test scores in elementary and secondary schools. This second year of our study therefore looks at test score trends since 2002.

To carry out this study, CEP collected, verified, and analyzed a vast array of state and national test data. On the state reading and mathematics tests used for NCLB accountability, we looked at two indicators of achievement—the percentages of students scoring at the "proficient" level, which is the main indicator of progress under NCLB, and a second indicator called effect size, which is based on average test scores. We also compared trends on state tests with trends on a second independent measure, the state-by-state results of the National Assessment of Educational Progress (NAEP), to see whether achievement had moved in the same direction and to the same degree on both assessments.

This year's study continues the focus on two main questions—whether reading and math achievement has increased since 2002 and whether achievement gaps between subgroups of students have narrowed. It updates our analyses of last year by including test results from 2007, the most recent year available at the time of our data collection, to help determine whether trends are being sustained over a longer term. It takes a more in-depth look at achievement gaps by including trends for specific grades at the elementary, middle, and high school levels, as well as trends across these three levels. In addition, this year's study includes more extensive comparisons of test score trends at grades 4 and 8 on state tests since 2002 and on NAEP between 2003 and 2007. Not only did we compare trends in percentages proficient on state tests with trends in the percentages of students scoring at or above the basic level on NAEP (the most appropriate comparison for reasons explained in chapter 4), but we also compared overall trends in effect sizes on the two assessments. In addition, we looked at trends in achievement gaps on both assessments.

The study was conducted with advice from a panel of five nationally known experts in educational testing or policy research and with technical support from our contractor, the Human Resources Research Organization (HumRRO). Five main conclusions emerged from our analyses. (Technical terms used in this report are defined in a glossary at the end of this chapter.)

Main Conclusions

1. Since 2002, reading and math achievement on state tests has gone up in most states according to the percentages of students scoring at the proficient level. Gains tended to be larger at the elementary and middle school grades than at the high school level. Achievement has also risen in most states according to effect sizes. These findings are drawn from states with at least three years of comparable test data.

2. Trends in reading and math achievement on NAEP have generally moved in the same positive direction as trends on state tests, although gains on NAEP tended to be smaller than those on state tests. The exception to the broad trend of rising scores on both assessments occurred in grade 8 reading, where fewer states showed gains on NAEP than on state tests, especially in terms of effect sizes.

3. In states with sufficient data to determine achievement gap trends on state tests, gaps have narrowed more often than they have widened since 2002, particularly for African American students and low-income students. Gap trends were also largely positive for Latino students, but this finding is less conclusive because in many states the Latino subgroup has changed significantly in size in recent years. On the whole, percentages proficient and effect sizes revealed similar trends of narrowing or widening, although percentages proficient gave a more positive picture of achievement gap trends than effect sizes.

4. Gaps on NAEP have also narrowed more often than they have widened in states with sufficient data to determine gap trends. The exception was in grade 8 math, where gaps on NAEP widened more often than they narrowed for most subgroups. In general, NAEP results painted a less positive picture of progress in narrowing gaps than state tests did.

5. It is impossible to determine the extent to which these trends in test results have occurred because of NCLB. Since 2002, many different but interconnected policies and programs have been undertaken to raise achievement—some initiated by states or school districts and others implemented in response to federal requirements. Moreover, all public school students have been affected by NCLB, so there is no suitable comparison group of students to show what would have happened without NCLB.

Findings That Support Main Conclusions

The conclusions above are based on specific findings about the numbers and percentages of states displaying various test score trends in the years since 2002 with comparable data. Highlights of these specific findings are listed below, and additional key findings appear at the beginning of each chapter of this report.

OVERALL ACHIEVEMENT

- **Number of gains and declines on state tests.** On the percentage proficient indicator, there were 133 instances of moderate-to-large gains on state tests and only 9 instances of moderate-to-large declines. On the effect size indicator, there were 121 instances of moderate-to-large gains and only 11 instances of moderate-to-large declines. (By "instance," we mean a trend in a particular state on one indicator for a specific grade and subject.)

- **Key trends in reading on state tests.** In elementary reading, 17 of the 28 states with both percentage proficient and effect size data posted moderate-to-large gains on both indicators, and 2 more states showed at least slight gains on both. No state showed a decline in percentages proficient, but 3 states exhibited declines in effect sizes. In middle school reading, 14 of the 28 states with sufficient data demonstrated moderate-to-large gains on both indicators, and 6 more states had at least a slight gain on both. Six states showed declines on at least one indicator in middle school reading. Of the 27 states with sufficient data in high school reading, 8 states made moderate-to-large gains on both indicators, and 7 more states made at least slight gains on both. High school reading achievement declined on both indicators at a moderate-to-large rate in 2 states and at a slight rate in 3 states. Other reading results, including trends in states with data on just one indicator, can be found in chapter 3.

- **Key trends in math on state tests.** In elementary school math, both percentages proficient and effect sizes increased at a moderate-to-large rate in 21 of the 27 states with sufficient data and increased at least slightly on both indicators in 5 more states. Only 1 state showed a decline of any size on either indicator in elementary math. In middle school math, 22 of the 27 states with sufficient data posted moderate-to-large gains on both indicators, 3 more states displayed at least slight gains on both indicators, and only 2 states experienced declines of any size on either indicator. At the high school level, math scores improved at a moderate-to-large rate on both indicators in 12 of the 26 states with sufficient data and declined on one or both indicators in 6 states. In 2 states, high school trends went up on one indicator and down on the other. More math results, including trends in states with data on just one indicator, are described in chapter 3.

- **Overall comparisons between state tests and NAEP.** At least 29 states had sufficient data to compare trends on state tests and NAEP (the number varied by subject and grade). Of the 134 comparisons made between state percentages proficient and NAEP percentages basic, gains occurred on *both* assessments in 108 cases and declines on both in just 2 cases. In the remaining 24 cases, trends on the two assessments diverged. Of the 118 comparisons made between effect size trends on state tests and NAEP, gains occurred on both assessments in 77 cases, no net change on both in 2 cases, and declines on both in no case. In the remaining 39 cases, trends differed. Trends on state tests and NAEP diverged most often in grade 8 reading, where percentages proficient/basic trends agreed 62% of the time, effect size trends agreed just 20% of the time, and NAEP results were less positive.

- **Smaller gains on NAEP.** Of the 252 comparisons made between the two assessments, state tests showed larger gains in 188 cases (75%), while NAEP showed larger gains in 55 cases (22%).

ACHIEVEMENT GAPS

- **Summary of gap trends on state tests.** In the states with sufficient data, we looked at achievement gaps in reading and math for four subgroups at three grade levels on two indicators. Gaps in percentages proficient narrowed in 327 instances, widened in 76 instances, and showed no net change in 20 instances. Gaps in effect sizes narrowed in 184 cases, widened in 56 cases, and showed no net change in 30 cases.

- **Trends for specific subgroups.** Gaps for the African American and low-income subgroups narrowed far more often than they widened in both reading and math. In elementary reading, for example, the African American-white gap narrowed in 13 states according to both percentages proficient and effect sizes and widened on both indicators in only 1 state. Gaps between low-income and non-low-income students narrowed in 10 states on both indicators, and in no state did this gap widen. At the middle and high school levels, there were also more cases of gaps narrowing than widening for the African American and low-income subgroups. More gap trends for these subgroups and for Latino and Native American students can be found in chapter 5.

- **Comparisons of gap trends between NAEP and state tests.** Of the 289 gap comparisons made between the two assessments, gaps narrowed on NAEP 62% of the time while they narrowed on state tests 80% of the time. Chapter 6 includes specific comparisons of gap trends on NAEP and state tests by subject, grade, and subgroup.

Possible Explanations for Trends Identified

This study of achievement was designed to present test data and identify test score trends rather than to investigate the reasons underlying those trends. However, we did draw from other research—most notably CEP's broader six-year study of NCLB implementation—to speculate about possible explanations for the trends we identified. Several factors, including the following factors alone or in combination, may explain why we found more evidence of test score gains than declines and of gaps narrowing rather than widening:

- Students may be learning more as a result of improvements in curriculum and instruction and extra interventions for urban students and for lower-performing students and subgroups.

- Many school districts are devoting more instructional time to reading and math—often by reducing time for other subjects.

- Students and teachers may have become more familiar with the format and content of state tests as a result of test preparation activities.

- Narrow types of test preparation may have led to "score inflation" on state tests, meaning that scores have gone up on state tests without students having learned the broader knowledge of the subject being tested.

- Subtle manipulations in test design not otherwise addressed in this study could have made some state tests easier over time. Examples include small changes in the choice of test items, training for scorers of open-ended test items, and procedures for equating scores across different test versions.

Several additional factors may account for some of the differences in trends on NAEP and state tests. First, state tests may better reflect what students are actually being taught in class than NAEP does. This is because state tests are designed to be aligned with a specific state's academic content standards, while NAEP is not deliberately aligned to any state's standards. Second, students may not be as motivated to perform well on NAEP as they are on state tests. In contrast to the high stakes attached to state test scores, NAEP results are not connected to any rewards, sanctions, or future outcomes for students or educators, and do not even include individual student scores. Third, because of data availability issues, our analyses sometimes looked at different grades or years of data on state tests and NAEP. These variations could yield somewhat different trends. Finally, differences in gap trends between state tests and NAEP may be partly attributable to small changes over time in the distribution of students' scores relative to the cut scores for proficient (or basic) performance and the mean, or average, scores on both assessments. More discussion about these and other possible explanations can be found in the final sections of chapters 3 through 6 of this report.

Noteworthy Features of the Study

Several features distinguish this study from other studies of student achievement. These features are described in more detail in chapter 2.

First, all 50 states have verified the accuracy of the test data used in this study.

Second, we have taken steps to minimize the effects of the natural year-to-year fluctuations in test scores that occur for reasons unrelated to learning and that could lead to potentially misleading conclusions based on a change in test scores over just one year. Specifically, the national analyses of achievement trends in this report include only those states with three or more years of comparable test results for a particular indicator, subject, grade, or subgroup. We have omitted test results affected by breaks in comparability, such as the adoption of a new test, a change in the cut score for proficiency, or other major changes in state testing systems. In addition, we have determined trends by looking at average yearly changes in achievement or achievement gaps across three or more years.

Third, we have taken a cautious approach in drawing conclusions about achievement for subgroups that are small or have changed significantly in size, as well as for students with disabilities and English language learners (ELLs). These two subgroups have been subject to shifting NCLB-related testing policies.

Fourth, we have developed a fixed set of criteria and applied them consistently across states to address the problem of how to form a national picture of student achievement from the results of 50 state testing systems that vary widely in content, format, difficulty, scoring scales, and proficiency definitions. Because of these variations, scores from different states' tests cannot, and should not, be compared directly with each other in the same way that one might compare different states' scores on the same national test. Recognizing this, we reached national conclusions by tallying the numbers of states with achievement gains and declines and with narrowing and widening gaps on their own tests. We also applied consistent criteria across states to determine what constituted a "slight" or "moderate-to-large" change in achievement and which years of data and which grades to analyze.

Availability of Test Data from States

In last year's report on achievement, we recommended improvements in the availability and transparency of state test data. States appear to have taken steps over the past year to improve data availability. Most notably, the number of states that provided the data needed to calculate effect sizes in at least some grades increased from 30 states last year to 42 this year, even though states are not required by NCLB to report this information. In addition, key data seemed to be easier to locate on state Web sites.

Even so, many states had fewer than three consecutive years of comparable test results that could be used to determine achievement trends for this study. In most cases, this was because states had adopted new tests during the past two years or had changed their cut scores. In light of the changes that states have recently made or plan to make in their testing systems, it seems likely that breaks in trends will be a reality for at least the next few years, complicating efforts to identify meaningful, long-term trends in some states.

Organization of This Report

Chapter 2 of the report provides background information about the study and describes its key features to help readers understand the data presented in the rest of the report. This chapter also discusses the availability of test data from the states.

Chapter 3 discusses our findings about overall achievement trends on state reading and mathematics tests since 2002. The chapter first looks at the direction of trends across both subjects and three grade levels (elementary, middle, and high school), and then examines separate trends in reading and math at grades 4, 8, and 10 or an adjacent grade in states where comparable data for these grades were not available. The chapter also considers the size of gains and declines in overall achievement. A final section offers possible explanations for the overall trends we found.

Chapter 4 compares the direction and magnitude of overall achievement trends on state tests and NAEP. We also offer possible explanations for the similarities and differences revealed by the two measures.

Chapter 5 examines achievement gaps for subgroups of students. We first discuss consistent trends across all three grade levels in the same subject, and then look more closely at specific trends by grade level, subject, and subgroup. We present enrollment data showing the concentration of particular subgroups in a limited number of states and consider possible explanations for the gap trends identified.

Chapter 6 compares trends in achievement gaps on state tests and NAEP. Both the direction of the trends and the size of the gaps are compared on the two assessments. The chapter also considers possible explanations for the similarities and differences in gap trends that emerged on state tests and NAEP.

Appendix 1 describes in more detail the methods used to conduct this study. Appendix 2 includes graphs that illustrate the concept of a standard deviation, which is important to understanding effect size, and explain why achievement gaps can appear larger or smaller depending on where the cut score is set on a test.

Profiles and Raw Data for Each State on the Web

The findings in this report are drawn from an extensive and in some ways unparalleled collection of test data from each state. These data have been assembled into a profile for each state. All 50 profiles have been posted on CEP's Web site at www.cep-dc.org. Readers who are mainly interested in one state or would like to learn more about trends in specific states are encouraged to go to the specific states' online profiles where they will find a wealth of data. The profiles include the following information:

- A summary of key findings about overall achievement and achievement gaps for that state in bullet form
- Two tables, one for reading and one for math, summarizing the overall achievement trends (up, down, or no change) and gap trends (narrowed, widened, or no change) at each grade level analyzed in the state
- Brief descriptions of data limitations in that state and key characteristics of the state's testing system
- Figures and tables with overall percentages proficient for all years with comparable data and for all grades tested under NCLB
- Figures and tables with overall effect sizes for all years with comparable data and for all grades analyzed for this study
- Tables with test results and achievement gap calculations for the state's main subgroups, shown first in terms of percentages proficient and next in terms of effect sizes
- A table with the number of test-takers in each subgroup for the starting year and ending year of the trends shown for that state
- Definitions of key terms used in the profile and caveats about the data

In addition, CEP has posted on its Web site all the raw test data collected from all 50 states for all grades tested for NCLB purposes, including grades not analyzed in this report.

Glossary of Key Terms

This report uses the following technical terms:

Percentage proficient. The percentage of students in a group who score at or above the cut score for "proficient" performance on the state test used to determine progress under NCLB.

Effect size. A statistic that conveys the amount of difference between test results using a common unit of measurement that does not depend on the scoring scale for a particular test. An effect size is computed by subtracting the mean scale score (see below) on a test for one year, such as 2006, from the mean scale score for another year, such as 2007, then dividing the result by the average standard deviation (see below).

Mean scale score. The arithmetical average of a group of test scores, expressed on a common scale for a particular state's test. The mean is calculated by adding the scores and dividing the sum by the number of scores.

Standard deviation. A measure of how much test scores tend to deviate from the mean—in other words, how spread out or bunched together test scores are. If students' scores are bunched together, with many scores close to the mean, then the standard deviation will be small. If scores are spread out, with many students scoring at the high or low ends of the scale, then the standard deviation will be large.

Moderate-to-large gain or decline. For percentages proficient, an average gain or decline of 1.0 or more percentage points per year. For effect sizes, an average gain or decline of 0.02 or more standard deviation units per year.

Slight gain or decline. For percentages proficient, an average gain or decline of less than 1.0 percentage point per year. For effect sizes, an average gain or decline of less than 0.02 standard deviation units per year.

Small subgroup. A racial, ethnic, or demographic subgroup composed of fewer than 500 test-takers in a particular grade across a state.

Changing subgroup. A subgroup in which the number of test-takers has increased or decreased by at least 25% over the period of years analyzed.

CHAPTER 2

Understanding the Achievement Data in This Report

Summary of Key Points

This chapter provides background information about this study and explains its noteworthy features to help readers understand the findings in the chapters that follow. The chapter also discusses the availability of test data from states. Key points from this chapter include the following:

- **This study is comprehensive, based on an extensive array of data from all 50 states and multiple types of test data.** Where available, we analyzed results on state tests from 2002, the year No Child Left Behind took effect, through 2007, the most recent year of data available when our research was completed. In addition to looking at percentages of students performing at the proficient level—the main indicator of progress under NCLB—we looked at effect size, a statistic based on average test scores that avoids some of the limitations of percentages proficient. To see whether improvements in state test scores also show up on an independent measure of achievement, we compared state test score trends with trends on the National Assessment of Educational Progress.

- **The achievement data in this study have been vetted to minimize the possibility of distorted or misleading trends.** Only trends of at least three consecutive years of state test data have been analyzed. Test results have been omitted if they were subject to "breaks" in comparability due to changes in that state's test or in cut scores for proficient performance. In addition, changes in test scores over time have been averaged to minimize the effect of natural year-to-year fluctuations. Finally, subgroups that are small, have changed significantly in size, or have been subject to major changes in NCLB-related testing policies have not been included in national analyses of achievement gap trends.

- **In this, or any other study of achievement, it is impossible to sort out the extent to which recent trends in achievement have occurred because of No Child Left Behind.** This is because states, school districts, and schools have simultaneously implemented a variety of different but interconnected policies to raise achievement since 2002. Moreover, all public school students have been affected by NCLB, so there is no suitable comparison group of students to show what would have happened without NCLB.

- **Tests are incomplete and imperfect measures of student learning.** Although test scores are the best means available to draw inferences about learning at the state and national level, they are an imperfect and incomplete measure of how much students have learned.

- **The availability of certain types of achievement data has improved over the past year.** However, many states lack comparable trend data for some analyses because of changes in their testing programs. Between 2007 and 2008, the number of states that made available the data needed to calculate effect sizes increased from 30 to 42. Moreover, crucial data seemed to be easier to locate on state Web sites. Still, many states had fewer than three years of comparable data for either percentages proficient or effect

sizes or both—too few years to discern a trend. Usually this occurred because the states had adopted new tests or changed their cut scores for proficient performance.

Background on the Study

For the second year in a row, the Center on Education Policy has conducted a study that looks at two major questions—whether reading and math achievement has increased since the No Child Left Behind Act took effect in 2002 and whether achievement gaps[1] between subgroups of students have narrowed.

This year's study builds on CEP's 2007 appraisal of student achievement described in the report, *Answering the Question That Matters Most: Has Student Achievement Increased Since No Child Left Behind?* (CEP, 2007). This year's study also draws on knowledge gained from CEP's broader, six-year study of state and local implementation of NCLB. As part of that broader study, CEP has conducted an annual survey of all 50 states, an annual survey of up to 349 school districts, and case studies in 43 school districts and dozens of schools. The broader study has explored a wide range of topics, such as the impact of the NCLB accountability requirements, challenges of meeting the law's achievement goals, and efforts to dramatically overhaul low-performing schools through the NCLB restructuring requirements. Findings from the broader study have been published in a series of annual reports, entitled *From the Capital to the Classroom*, and in more than 30 special-topic reports, all available at www.cep-dc.org.

This year's achievement study incorporates several new features and improvements, including the addition of test results from 2007; more detailed analyses of achievement gaps at the elementary, middle, and high school levels; and more in-depth comparisons of trends on state tests and NAEP.

The test data analyzed for this report extend through 2007 in most cases and represent the most recent cycle of test results reported for NCLB by the time our data collection ended in April 2008. The 2007 data included in this year's report come from tests administered in spring 2007 or fall 2006 in states with fall testing. (Although states have since completed another cycle of testing during school year 2007-08, results from those tests typically are not available until the summer of 2008 or later.) With the addition of the 2007 data, we can begin to see whether the test score gains described in last year's achievement report are being sustained over a longer term.

This study is especially timely because the Congress and the new President will need to make decisions next year about the future of the No Child Left Behind Act and the appropriate role of the federal government in improving education. Policymakers across the political spectrum will benefit from having current information from all 50 states about test score trends since 2002. Moreover, because 2008 is the midway point between 2002-03 and 2013-14, the deadline for meeting NCLB's ultimate goal of 100% of students reaching the proficient level or above, this year represents a propitious time to assess progress in raising test scores.

[1] An achievement "gap" is the difference in test results between two groups of students, including various racial, ethnic, and demographic subgroups. For example, if 61% of Latino students and 83% of white students scored at the proficient level on a state test, the Latino-white gap would be 22 percentage points.

The study has been carried out with extensive technical support from our contractor, HumRRO, and with advice from a panel of five nationally known experts in educational testing or education policy:

- Laura Hamilton, senior behavioral scientist, RAND Corporation
- Eric Hanushek, senior fellow, Hoover Institution
- Frederick Hess, director of education policy studies, American Enterprise Institute
- Robert L. Linn, professor emeritus, University of Colorado
- W. James Popham, professor emeritus, University of California, Los Angeles

The panel provided advice on aspects of the study design and implementation, reviewed data, and reviewed and commented on drafts of this report. Although the panel members, as well as HumRRO staff, provided input on this report, we did not ask them to endorse it, so the findings and views expressed here are those of CEP.

Noteworthy Features of the Study

Below we describe the main features of this study. We have provided this information because we recognize that achievement trends are a serious matter requiring careful study and because some readers may be interested in the methods we used to collect and analyze the data in the chapters that follow. The explanations in this section have been grouped by topic and presented in a question and answer format so that readers can skip to the topics they want to learn more about. Topics covered include the following:

- Number of participating states
- Measures of achievement used, including their advantages and limitations
- Steps we took to minimize distortion of trends
- Criteria for determining which years and grades to include in trends
- Criteria for determining whether changes in test results were moderate-to-large, slight, or showed no change
- Policies for analyzing data for subgroups
- Approach for drawing national conclusions from 50 different state testing systems.

Additional details about our study methods can be found in appendix 1 at the end of this report.

HOW MANY STATES PARTICIPATED?

All 50 states participated in this study, as they did last year, making this the most comprehensive NCLB-related study of its type. (The District of Columbia chose not to participate in either year of the study.) The success of this study depended on the involvement of the states. We greatly appreciate the cooperation we received from state education personnel in all states.

Rather than assuming that the test data reported on state Web sites were accurate, we gathered the relevant data and asked all states to verify their accuracy. We gave states an opportunity to fill in missing information and make corrections in their current and prior years' data. Over two years, this comprehensive approach has yielded an extensive and in some ways unparalleled array of achievement data from every state—a possible maximum of 18,000 individual numbers per state. As discussed later in this chapter, however, not all states had sufficient data for every type of analysis described in this report.

WHICH MEASURES OF ACHIEVEMENT DID WE USE AND WHAT ARE THEIR ADVANTAGES AND LIMITATIONS?

The main measure of student achievement for this study consists of data from the state tests in reading (or English language arts at the upper grades) and mathematics used for accountability under the No Child Left Behind Act. Although no large-scale test provides a complete picture of achievement, we have focused on state test results because these tests are given to nearly all students in a state, are intended to reflect that state's academic content standards, and are designed to assess whether students have met their state's expectations for performance at a particular grade level.

This year we have given equal and independent weight to two main indicators of achievement from state tests: the percentages of students performing at or above the proficient level and, where available, a statistic called effect size. As explained below, each of these indicators has advantages and limitations. When used together they can compensate for each other's limitations and provide a fuller picture of achievement than either indicator could do alone.

This year we have also done more extensive analyses using a another measure of achievement, the National Assessment of Educational Progress—the only national assessment of what U.S. students know and can do in core academic subjects. By comparing trends on state tests and NAEP, we have tried to gauge whether students have made general improvements in learning that show up on more than one assessment. In particular, we have compared trends in percentages proficient on state tests with trends in the percentages of students reaching the "basic" level of performance on NAEP at grades 4 and 8. (Comparisons with grade 12 NAEP are not included because results at this grade are not broken down by state.) We have also compared progress on narrowing achievement gaps on state tests and NAEP. More information about the reasons for using the NAEP basic level, the differences between NAEP and state tests, and the methods used to compare results are discussed in chapters 4 and 6.

Percentages Proficient and Their Advantages and Limitations

Under NCLB, the progress of schools and districts in raising student achievement is determined largely by the percentages of students scoring at or above the proficient level on state tests. Percentages proficient are the only type of achievement data that states are required to collect and report under NCLB and, as such, have become a familiar indicator of progress in news stories, research studies, and state and school district "report cards" to parents.

Although percentages proficient are designed to indicate how many students have met a state's performance expectations, they have several limitations, including the following:

- "Proficient" means different things in different states. Each state sets its own cut score that defines proficient performance on its tests, and these definitions of proficiency vary considerably in ways that are sometimes murky. (See figure 6-A in chapter 6 for examples of how widely definitions of proficiency vary among the states.) Moreover, state tests

vary in their difficulty, content, and other aspects. One state may have higher percentages of students reaching proficiency than another state not because its students are learning more but because its tests are easier or its cut scores for proficiency are lower.

- Percentages proficient do not capture changes in student performance below or above the proficient cut score, such as an increase in the number of students moving from "below basic" performance to the "basic" level, from the proficient to the "advanced" level, or from barely proficient to almost advanced.

- The size of the achievement gaps between various subgroups depends partly on where a state sets its cut score for proficient performance. If a cut score is set very low or very high on the scoring scale for a test, so that almost everyone reaches it or almost no one reaches it, there is little apparent gap. But if the cut score is set closer to the middle of the scale where substantially more students have scored, the gaps between subgroups will appear larger. Appendix 2 explains this phenomenon in more detail.

Effect Sizes and Their Advantages and Limitations

To avoid some of the limitations of percentages proficient and provide an additional perspective on achievement, this study also analyzed an indicator called effect size. Effect sizes are computed from two types of statistics, both of which were collected for this study:[2]

a) *Mean test scores*, a type of average score determined by adding up all the individual scores of students on the numerical scoring scale for a particular state's test and then dividing the sum by the number of test-takers in the group.

b) *Standard deviations*, which are statistics that provide information about the distribution of test scores—in other words, how spread out or bunched together test scores are around the mean of a set of test scores. If test scores are bunched together, so that nearly all students score close to the mean, then the standard deviation will be small; if test scores are spread out, meaning that many students score far from the mean, then the standard deviation will be large. Appendix 2 includes a graph explaining the concept of standard deviations.

In essence, effect sizes are a measure of change compared to a standard deviation. An effect size is computed by subtracting the year 1 mean test score from the year 2 mean test score and dividing by the average standard deviation of the two years. Where there has been no change in average test score, the effect size is 0. An effect size of +1 indicates a shift upward of 1 standard deviation from the previous year's mean test score (although effect sizes tend to be much smaller than 1 for mean changes from year to year). Even if two states have very different scoring scales and proficiency cut scores, effect sizes can be used to make limited comparisons of changes in test scores between states or groups of test-takers in terms of standard deviation units.

What does it signify when one says, for example, that the average reading score for 4th graders increased by 0.19 of a standard deviation between 2002 and 2007? One way to interpret this would be to think of the change in terms of percentile ranks. A student who scores at the 50th percentile—in other words, higher than half of the students—has a percentile rank of 50. One standard deviation above the mean corresponds to a percentile rank of 84, assuming a normal distribution of scores. If the average test score in a state improved by one full standard deviation between 2002 and 2007, this would constitute a huge leap in

[2] For this study, we computed a standardized measure of effect size called Cohen's D (Cohen, 1992).

student performance—the equivalent of an increase from the 50th percentile to the 84th percentile. If a state increased student performance by 0.19 (about one-fifth) of a standard deviation, that is still a respectable gain.

Effect sizes are not susceptible to some of the same limitations as percentages proficient. Unlike percentages proficient, effect sizes provide a standardized index for measuring achievement gaps between subgroups that is not dependent on the proficiency cut score. For this reason, we were able to calculate effect sizes when the proficiency cut score had been changed on an otherwise unaltered test. Effect sizes also give a more complete picture of the entire range of student performance, both above and below the proficient level. If, for example, test scores went up for the lowest-achieving or highest-achieving students—those scoring well below or well above the proficient cut score—the effect size could increase while the percentage proficient decreased. Or, if a state focused intensively on raising achievement for students who are close to proficiency and might clear the bar with a little extra instruction while devoting relatively less attention to other students, the effect size could drop while the percentage proficient climbs.

Effect sizes have their own limitations, however. Effect sizes are more difficult to interpret than percentages proficient. Furthermore, effect sizes measure growth relative to other students' performance rather than to a set standard, so they do not address the critical question in many parents' minds of whether students have met the performance expectations for their grade level.

A Word about the Meaning of "Gains" under NCLB

This study determined achievement gains by comparing test results for one year's class of students, such as the 4th grade group in 2007, with results from a previous year's class, such as the 4th group in 2002, rather than by tracking the achievement gains of individual students as they progressed from 4th grade in 2002 to 9th grade in 2007. This model was used because it is consistent with how most states report achievement for NCLB purposes.

Difficulty of Attributing Causes and General Limitations of Tests

We tried to address some factors that can undermine efforts to reach accurate conclusions about achievement, but some limitations could not be addressed. These include the difficulty of attributing causes for the achievement trends we have identified and the general limitations of tests as a measure of achievement.

Although this report describes trends in test scores that have occurred since the enactment of NCLB, it is impossible to sort out the extent to which these trends have occurred because of NCLB. This is because many federal, state, and local reforms have been implemented at the same time as NCLB, and they are connected in ways that cannot be disentangled. Furthermore, there is no suitable "control group" of public school students not affected by NCLB that can be used for comparison purposes.

Readers should also keep in mind that test scores are not synonymous with achievement. Tests are imperfect and incomplete measures of how much students have learned. Only two subjects, reading and mathematics, are currently tested in a form that allows the type of analyses we conducted for this study. Even within those subjects, not all of the important knowledge and skills are measured by large-scale accountability tests. In addition, test scores can drift upward over time without actually indicating that students have learned more. Such increases in test scores may be due to growing familiarity with a test or certain forms of teaching to the test. Nevertheless, tests are widely regarded as the most defensible way to draw inferences about student learning at the state and national scale.

HOW DID WE ADDRESS FACTORS THAT COULD DISTORT TEST SCORE TRENDS?

We took steps, and applied them consistently across states, to minimize the effects of the natural fluctuations in test scores that occur for reasons unrelated to student learning and to factors that could produce distorted or misleading test results. The approach we took was intended to present as much information about achievement as possible while still maintaining a reasonable degree of confidence that the reported trends were accurate.

Limiting Trends to Years with Comparable Data

Many states have adopted changes in their testing systems in recent years that have created a break in test data, meaning that test results after the change should not be compared with results from before the change. For example, test results should not be compared if a state has introduced a new test (and has not taken steps to equate the scores on the old and new tests). Breaks in data can also occur if a state has changed the cut score for proficiency on the test, adopted a new scoring scale, or switched from a spring to a fall testing schedule. Ignoring these data breaks can produce a misleading picture of achievement.

To identify when a trend line had been broken, we gathered information from states about changes in their testing programs. We then limited our analyses of trends to only those years with comparable test data.

Averaging Trends over Three or More Years

Changes in test scores do not always move up or down in a straight line. More often, scores zigzag a bit from year to year, even if they are generally moving in one main direction. The figures below illustrate patterns of fluctuations that are somewhat typical in state test results. **Figure 2-A** represents a fairly gentle rise with some year-to-year variations, while **figure 2-B** exhibits a somewhat greater degree of fluctuation.

Figure 2-A. Percentage of Mississippi Students Scoring at the Proficient Level or Above in Mathematics

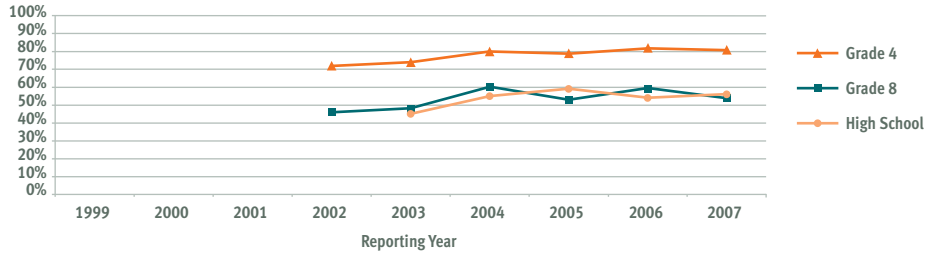
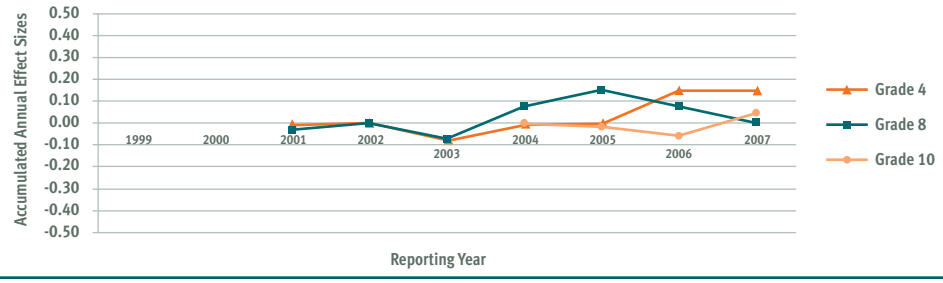


Figure 2-B. Reading Achievement Trends in South Carolina in Terms of Effect Sizes



Some of this year-to-year fluctuation in test scores may be due to factors unrelated to student learning that are more subtle than the adoption of a new test or a change in cut score. Often these factors cannot be addressed directly because they are not apparent or known.

A main reason for fluctuations is equating error, which indicates that one year's version of a test is not exactly comparable in difficulty to a previous year's version of the same test. (States often issue new versions of the same test to prevent cheating and excessive familiarity with specific test questions.) Test developers generally attempt to make new versions of a test as similar as possible to previous versions in terms of content and difficulty, but differences may still remain. Equating procedures are used to minimize the impact of these small differences. Still, a degree of uncertainty always remains because the new and old test scores are not perfectly equivalent; this is referred to as equating error. Other sources of variations between test versions include differences in the mixture of various types of test questions and their relative weights, and the guidelines and training given to people who score short-answer and essay questions.

A second reason for test score fluctuations unrelated to learning is a demographic change in the test-taking population. This could result from such factors as an influx of refugees, many families being uprooted due to a natural disaster, or a large number of students being held back as a result of new grade promotion policies.

Because of these sorts of fluctuations, it can be misleading to base judgments about achievement on a single year of test results or to draw conclusions about the direction of a trend from the difference between two data points. This study used two techniques to minimize the effects of these types of fluctuations and provide a more accurate picture of achievement over time.

First, we considered a movement in test results to be a "trend" only if it was based on three or more consecutive years of data—for example, percentages proficient on tests administered in school years 2004-05, 2005-06, and 2006-07. For each specific subject and grade combination or each particular subgroup, states that had just one or two years of data were omitted from the national summary counts of trends in chapters 3 through 6 and were categorized as having "too few years" of data or "data unavailable." However, data for these states have been included in the online state profiles developed for this report and posted at www.cep-dc.org. In these profiles, shifts in test results in states with just two years of data are referred to as "changes" in achievement rather than trends.

Second, we based our determinations of trends on average yearly gains or declines calculated from three or more consecutive years of data. Using averages helped to even out year-to-year fluctuations and provided a common metric of progress in all states analyzed, even when the test results sometimes covered different spans of years. Where only one or two years of comparable data were available, we did not calculate an average rate of change; in fact, it is not possible to calculate an average with one year, nor can the difference between two years be considered an average.

HOW DID WE DETERMINE WHICH YEARS OF TEST DATA TO ANALYZE?

Because we were focusing on trends across three or more years, the specific years for which states had comparable data varied. When comparable data were available, we analyzed state test score trends from 2002 through 2007. When states did not have comparable test data for this entire period due to the breaks described above, we used other starting or ending years within this span that would yield at least three years of comparable data. The specific years with comparable data for every grade are shown in the state profiles. The profiles also include test data going back to 1999 in states where these data are comparable to post-NCLB test results.

If a state introduced a new test in 2006, we did not include data for this state in the national summaries of trends in chapters 3 through 6. However, we did include data for 2006 and 2007 in the online state profiles for these states, referring to any shift in achievement over the two years as a change rather than a trend. When a new test was introduced in 2007, we used data from 2002 through 2006, or whatever period within this span that provided three years of comparable data.

The years with comparable data sometimes differed for the percentage proficient and effect size indicators. A small number of states provided fewer years of effect size data than of percentage proficient data. In some states, the years with comparable data also varied by subject or grade. This occurred when states had changed their tests in one subject or in selected grades—introducing a new math test, for example, while keeping the reading test the same, or instituting new high school exams while maintaining tests for grades 3 through 8—or when states phased in testing at additional grades to comply with NCLB testing requirements.

HOW DID WE DECIDE WHICH GRADES TO ANALYZE?

Within the time frame for this study, it would be almost impossible to analyze the overall achievement trends and gap trends in every state for every subgroup at each of the seven grades tested for NCLB (grades 3-8, plus one high school grade). To develop a reasonably rich picture of achievement trends in a realistic amount of time, we selected three grades as the targets of our main analyses—one at the elementary level, one at the middle school level, and one at the high school level. Grades 4, 8, and 10 were the default choices, not only because most states had sufficient data at these grades but also because these choices would allow us to make the most relevant comparisons with the state-by-state NAEP data for grades 4 and 8. If a state did not have three years of comparable data at these default grades, an adjacent grade with sufficient data was chosen in the same fixed order for every state, as explained in appendix 1.

In some states, the grades analyzed have changed since last year's study. This occurred when three years of data became available at the default grade or when a previous trend line was broken by the introduction of a new test in a particular grade.

WHAT DO MODERATE-TO-LARGE, SLIGHT, AND NO CHANGE MEAN?

If the amount of change in test scores is very small, this change should be viewed more cautiously than larger changes because small changes may reflect natural fluctuations in test scores rather than real changes in achievement. For this reason, we established criteria for characterizing gains or declines as "slight" or "moderate-to-large." In consultation with the expert panel, we defined a slight gain or decline as an *average yearly change* in the percentage proficient of less than 1.0 point per year, or an average yearly change in effect size of less than 0.02 standard deviation units. Changes of 1.0 percentage point per year or more, or 0.02 standard deviation units per year or more, were considered moderate-to-large. By applying these definitions, we could determine the magnitude of changes in achievement on either indicator.

When the differences in percentages proficient averaged across three or more years amounted to 0.0 or when the differences in effect sizes averaged 0.00, this was considered as "no change" in achievement. Although slight ups and downs may have occurred in the interim years, there was no net gain or decline in these situations.

One reason we report slight gains in this study is that over a period of several years, consistent slight gains can add up to a respectable amount of improvement. For example, average yearly gains in effect size of 0.02 over 12 consecutive years would amount to almost a quarter of a standard deviation, or a move from roughly 50% to 60% proficient.

While a 1 percentage point difference is a relatively familiar concept—the difference, say, between 69% proficient and 70% proficient—a difference in effect size of 0.02 is more difficult to grasp. One can get an idea of the magnitude of effect size gains from other studies that have used this measure:

- In studies comparing the international ranking of U.S. students in mathematics with those of other countries, an effect size increase of 0.7 standard deviations in U.S. math scores would raise the U.S. from the middle of the pack of 41 countries to one of the top five (Black & Wiliam, 1998). An upward move of about one quarter of a standard deviation would halve the achievement gap in math between U.S. students and those in some top-performing Asian countries.

- An achievement gap of 1 standard deviation is roughly equivalent to the gap in math scores between U.S. students and Japanese students (Peterson & West, 2003).

- Between 1967 and 1982, scores of U.S. students on the SAT college admissions test fell by 0.3 standard deviations. Between 1970 and 1982, high school science scores on NAEP fell by 0.4 standard deviations. At the time, these drops were considered by some researchers to be alarming (Peterson & West, 2003).

- Another study (Levitt & Fryer, 2004) found that, on average, African American students score about 1 standard deviation below white students on standardized tests.

WHAT CRITERIA DID WE USE TO ANALYZE TRENDS FOR SUBGROUPS OF STUDENTS?

In this study, we took two steps to minimize distortions in trends for subgroups of students. First, we took care not to draw conclusions about achievement trends for subgroups that were small or had changed significantly in size. Second, we did not include trends in our national counts for students with disabilities and English language learners, two subgroups that have been subject to shifts in NCLB-related testing policies.

Small or Changing Subgroups

When the number of test-takers in a group is small, the score of each student has a greater impact on the performance of the whole group. Test results for a small group are often less stable and reliable than those of larger groups, and test score changes may be more a reflection of changes in group composition than of changes in student learning.

Similarly, rapid changes in the size of a group can also affect test results. For example, if a school district has experienced an influx of immigrants from Latin American or Asia, the Latino or Asian subgroup could change from a small group of students who were born in the U.S. or had lived here for a long time to a larger group of new arrivals with little or no English language skills and a substandard education in their native country. In other words, the Latino or Asian students who took the state test in 2007 could include a very different population from those who took the test in 2004, and this could affect overall test results for that subgroup. In this situation, one could not fairly assume that a decline in test scores means that Latino 4th graders are being taught less effectively or have learned less than their 2004 counterparts.

Several states have small racial or ethnic subgroups because they enroll relatively few students of color. In addition, many states have experienced substantial growth or declines in the size of some subgroups, such as rapid growth in Latino enrollments where there had been few such students before.

To address these issues, we developed definitions of small or changing subgroups for this study based on guidance from the expert panel. A small subgroup is one that consists of fewer than 500 test-takers at a particular grade across a whole state. A changing subgroup is one in which the number of test-takers has increased or decreased by at least 25% during the years analyzed. In the main analyses of achievement gap trends in chapter 5, we have omitted small and changing subgroups. But we have also included separate tables showing gap trends in states where a specific subgroup is small or changing. In the state-by-state summary table at the end of chapter 5, as well as in the online state profiles accompanying this report, we have flagged with footnotes the subgroups in a particular state that are small or changing. These footnotes urge users of the reported data to exercise caution in drawing conclusions.

Students with Disabilities and English Language Learners

A different but related set of issues affects the interpretation of test results for students with disabilities and English language learners (also known as limited English proficient students). Federal and state policies for testing these two subgroups have changed in notable ways since 2002. These changes have affected which students in these subgroups are tested for accountability purposes, how they are tested, and when their test scores are counted as proficient under NCLB.

For students with disabilities, the federal government and most states have refined their policies over the past four years to clarify which students can take tests geared to "alternate" standards based on their learning level rather than their grade level, and which students can take tests with "modified" standards. Federal regulations have also clarified when scores from assessments based on alternate and modified standards can be counted as proficient for NCLB purposes. To implement these policy changes, many states have recently introduced new types of assessments for students with disabilities.

For English language learners, changes in federal guidance have clarified that ELLs in their first year of attendance in a U.S. school may be exempted from taking the test used for NCLB accountability. In addition, states may allow students who were formerly English language learners but have achieved English language proficiency or fluency within the past two years to be counted in the ELL subgroup for NCLB accountability. Furthermore, several states have introduced new versions of tests aimed at ELLs in recent years, such as tests that use a less complex English vocabulary to assess subject-area knowledge.

As a result of these policy revisions, some students with disabilities and ELLs who were excluded from testing in the early years of NCLB may now be included, while some who had been included earlier may now be exempted or tested differently. It has taken time for states, districts, and schools to figure out which students in these subgroups should be tested with regular tests, with accommodations in test administration (such as being given more time to complete the test, taking the test in a small group setting, or using a large-print test version or a dictionary), or with tests aligned to alternate or modified standards. All of these changes could affect the year-to-year comparability of test results for these subgroups. For this reason, the analyses of achievement gaps in chapters 5 and 6 do not include national trends for students with disabilities and English language learners. Test results for these subgroups do appear online in the state profiles with a note suggesting that caution be used when interpreting trends.

A Note about the Asian Subgroup

The Asian subgroup was not included in our national analyses of gap trends because in most states, Asian students as a group performed as well as or better than white students. Test data for Asian students are shown in the individual state profiles, however.

HOW DID WE DRAW NATIONAL CONCLUSIONS FROM 50 STATE TESTING PROGRAMS?

As noted above, state tests vary greatly in their content, format, difficulty, scoring scales, and proficiency definitions. For that reason, percentages proficient or other scores from different states' tests should not be compared directly with each other in the same way that one would compare scores from the same test. How, then, is it possible to form a national picture of student achievement trends from 50 disparate testing systems? The general approach we used was to determine the direction and size of achievement trends in each state—whether overall achievement had increased or decreased and to what degree, and whether gaps had narrowed or widened—and then to count the number of states with gains and declines or with narrowing or widening gaps (or in a few cases with no net change over time). We used a similar approach to determine whether trends had moved in the same direction on state tests and NAEP.

To make these determinations of trends, we analyzed the test data provided by the states and included in each state's online profile. We used a rigorous process in which three senior analysts—a CEP consultant, a CEP staff person, and a HumRRO staff person—independently coded the overall achievement trends and gap trends for each state. The three analysts noted whether overall achievement had increased slightly, increased to a moderate-to-large degree, decreased slightly, decreased to a moderate-to-large degree, or showed no net change. They also noted whether achievement gaps had narrowed, showed no net change, or widened. The analysts then checked with each other to make sure that their coding of trends agreed and reviewed the state data if their coding did not correspond. When the coding process was completed, they tallied the number of states with each type of trend, again checking to make sure their tallies agreed.

Many of the tables in this report display counts of states showing different combinations of trends in a grid format. **Box 2-A** explains how to read these grids, which appear throughout the report.

Box 2-A. How to Interpret the Data in the Grid Tables

Many of the tables in this report use a grid format, like the one in **table 2-A**. These grids convey certain information in a continuum that cannot be easily captured with a traditional table or bar graph. Each row and each column shows a continuum of trends, moving from a moderate-to-large gain on the left side of a row (or the top of a column) to a moderate-to-large decline on the right side of a row (or the bottom of a column). As the shaded bands at the top and left sides of the table indicate, the rows across the table follow this continuum for the percentage proficient indicator, while the columns down follow this continuum for the effect size indicator. We have also assigned symbols (arrows, stars, or diamonds) to each type of trend.

Each cell in the table represents an intersection of a row (a percentage proficient trend) and a column (an effect size trend). For example, the unshaded cell in the upper left, which includes 17 state abbreviations, shows the number of states that had a moderate-to-large gain in percentage proficient (the first symbol in the cell) and also had a moderate-to-large gain in effect size (the second symbol in the cell). Moving toward the right across that row, the next cell that is not empty indicates that one state, Oregon, had a moderate-to-

Continued on page 21

Box 2-A (continued)

large gain in percentage proficient and no change in effect size. The next cell to the right shows that one state, New Mexico, had a moderate-to-large gain in percentage proficient but a slight decline in effect size. The cell on the far right of the row reveals that five states had moderate-to-large gains in percentages proficient but did not have sufficient comparable data to determine an effect size trend.

Table 2-A. Test Score Trends Since 2002 in Elementary Reading

Percentage of students scoring at or above the proficient level	Effect size						
	Type of Trend	↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	♦/❍ Data unavailable or too few years
	↑ Moderate-to-large gain	↑ ↑ 17 states: AR, CA, FL, ID, IA, KY, LA, MS, NV, ND, OH, OK, SC, TN, TX, WA, WV		↑ ✪ 1 state: OR	↑ ↘ 1 state: NM		↑ ♦/❍ 5 states: AL, AK, MD, MT, NE
	↗ Slight gain	↗ ↑ 2 states: NC, UT		↗ ✪ 5 states: CO, HI, NJ, PA, SD	↗ ↘ 1 state: IN	↗ ↓ 1 state: AZ	↗ ♦/❍ 1 state: MA
	✪ No change						
	↘ Slight decline						
	↓ Moderate-to-large decline						
	♦/❍ Data unavailable or too few years	♦/❍ ↑ 1 state: DE					♦/❍ ♦/❍ 15 states: CT, GA, IL, KS, ME, MI, MN, MO, NH, NY, RI, VT, VA, WI, WY

Similar grids are used in chapter 5 to compare trends in achievement gaps. In these grids, the row and column headings indicate whether a gap for a specific subgroup narrowed, showed no change, or widened in terms of percentages proficient (rows across) and effect sizes (columns down). Grids are also employed in chapters 4 and 6 to compare trends on state tests and NAEP. In these comparisons, the state test score trends appear across a row, and the NAEP trends appear down a column. The cells of the grid indicate how many states had various combinations of trends on state tests and NAEP.

Availability of State Data for This Study

In last year's report on achievement, we noted that our efforts to conduct a rigorous study were hampered by incomplete and inconsistent data and by a lack of data needed to compute effect sizes in many states. We also suggested that transparency in state test data could be improved to make achievement data more readily available to the public and easier to understand.

This year we found that the availability of state test data had improved, particularly effect size data. All states provided percentage proficient data for our study. And the number of states that provided the data needed to calculate effect sizes in at least some grades increased from 30 last year to 42 this year. NCLB does not require states to report the mean scale scores and standard deviations needed to compute effect sizes. In the past, these data have not been readily, or at least publicly, available in several states, and states may have had few requests for them. This year, however, it appears that more states have recognized the value of releasing this type of test information to researchers and the public.

Compared with last year, data were also somewhat easier to locate on state department of education Web sites. This may be partly because HumRRO's data collection procedures improved this year, but it also appears that some states have made greater efforts to post direct links to their state achievement report cards on their home pages. In fact, several state education department personnel mentioned to HumRRO staff that the state had made specific improvements to data availability after taking part in this study last year.

We also experienced greater cooperation from state department of education personnel in obtaining data that could not be located on the Web. On the whole, states seemed better prepared to provide the necessary data, especially when the same person was the state point of contact last year and this year.

Although more states provided a variety of test data this year, not all states had comparable data for the three years needed to discern a trend. Many states had fewer than three years of comparable data for either the percentage proficient indicator or the effect size indicator or both, but the number varied depending on subject, grade, or subgroup. As noted above, this was generally because the states had made a major change in their testing program and had just one or two years of comparable test results. The states that lacked sufficient comparable data this year were not always the same ones that lacked data last year. If a state introduced a new test in 2005-06, for example, this year we would have reported two years of results from 2006 and 2007, the most recent test administrations but too few years to consider a trend. Last year for that same state we might have reported results from the old test for 2002 through 2005, enough years to consider a trend. (See appendix 1 for a more thorough explanation.)

As part of this study, we gathered information from states about major changes in their testing programs since 2002. These reports indicate that the lack of comparable data is due to three main types of changes in testing policies (with some states making more than one of these changes):

- Fourteen states introduced new tests or new versions of tests in school year 2005-06 or 2006-07.
- Six states revised their scoring scales, cut scores, or performance standards in ways that affected score comparability.

- Two states shifted the time window for administering their tests from winter to fall or from fall to spring, a move that affected comparability with previous scores because students were being tested at a different point the school year.

Subgroup data was spottier than overall achievement data in some states. A few states provided effect size data for students as a whole but had not disaggregated these data by subgroup. In some states, data were missing for certain subgroups in particular grades or years. Collecting comparable data for English language learners seemed to be especially difficult, most likely because of policy changes affecting which students were included in that subgroup in recent years.

Overall, state reports about their testing systems suggest that the testing landscape is continually changing, as states adopt new tests and phase out old ones, revisit content and performance standards and redesign tests around new standards, lower or raise cut scores, and develop assessments for students with disabilities and English language learners. In this atmosphere, breaks in trend data are likely to be a reality in some states for at least the next few years.

CHAPTER 3

Trends in Overall Achievement on State Tests

Summary of Findings

In this chapter, we look at the question of whether reading and mathematics achievement has increased on state tests since 2002 for students overall. The results of our analyses can be summarized as follows:

- **The number of instances of gains on state tests far outnumbered declines.** For all of the states with sufficient data, we tallied trends at each grade level analyzed in both reading and math on two indicators. On the percentage proficient indicator, there were 133 instances of moderate-to-large gains, compared with only 9 instances of declines of similar magnitude. On the effect size indicator, there were 121 moderate-to-large gains and only 11 moderate-to-large declines.

- **In reading, the number of states showing gains in achievement since 2002 far outstripped the number showing declines at all three grade levels analyzed.** Trends were most positive at the elementary level. Gains tended to be slighter at the middle and high school levels than at the elementary level. More states experienced declines in achievement at the high school level than at the other grade levels, although gains still outnumbered declines.

- **In reading, analyses of percentages proficient and analyses of effect sizes generally yielded the same trends, but in a few states the two indicators produced contradictory results.** In elementary school reading, something of a pattern emerged, with percentages proficient giving a more positive picture than effect sizes in some states.

- **In elementary and middle school mathematics, the trends were overwhelmingly positive.** In almost all states with sufficient data, achievement increased according to both percentages proficient and effect sizes. Gains were also evident in most states with data for just one of these indicators. Elementary school mathematics was the area in which the largest number of states showed improvement. At the high school level, many more states experienced declines in achievement than at other grade levels, although states with gains still outnumbered those with declines.

- **In math, achievement trends based on percentages proficient were quite consistent with trends based on effect sizes.** In only a few cases did the trend go up on one indicator and down on the other.

- **Gains tended to be smaller at the high school level than at the elementary or middle school levels.** This pattern of smaller average yearly gains in high school was apparent for both percentages proficient and effect sizes and in both reading and mathematics.

How We Analyzed Overall Trends in State Test Scores

We examined the directions of trends in state test scores in two subjects (reading and math), at three grade levels (one grade each at the elementary, middle, and high school levels), according to two indicators (percentages proficient and effect sizes). This amounts to 12 possible trends in each state. To determine a trend, we needed at least three consecutive, comparable years of test data. We computed average yearly gains or declines in both percentages proficient and effect sizes. These were determined by subtracting each state's starting results from its ending results and dividing by the number of years in the trend line. For example, Arkansas had high school test data from 2002 to 2007, which meets the trend criterion of at least three years of data. In reading, 37% of Arkansas high school students were proficient in reading in 2002; this increased to 51% in 2007. The difference between the two figures of 14 percentage points was divided by five years to produce an average yearly gain of 2.8 percentage points. A similar method was used to determine average yearly effect sizes.

We coded each trend as a moderate-to-large gain, slight gain, no net change, slight decline, or moderate-to-large decline. As explained in chapter 2, this study defined a "moderate-to-large" gain or decline as an average yearly change of at least 1.0 percentage point per year in the percentage proficient, or at least 0.02 standard deviation units per year in effect size. Gains or declines below these thresholds are considered "slight."

To get an initial idea of the nationwide picture, we tallied trends for all of the states with sufficient data, counting a trend in each state for a specific grade, subject, and indicator as an "instance." For example, in Colorado between 2002 and 2007, the percentage proficient increased slightly in reading at the elementary grade analyzed. This is counted as one instance of a slight increase. Each cell in **table 3-I**, a state-by-state summary table that appears later in this chapter, represents a possible instance.

We looked at how many states showed consistent gains or declines across all three grade levels. Next, we analyzed separate trends in reading and math at each of the grade levels, noting which states made gains. We then examined the size of gains and declines in both subjects. Finally, we considered possible explanations for the trends we found.

Direction of Trends in Test Scores Nationwide

We sought to determine whether state test scores had generally gone up, down, or stayed the same since 2002. To get the overall picture, we simply counted the number of instances of various trends (moderate-to-large, slight, and no change) in table 3-I.

The results of our national tally are displayed in **table 3-A**. The number of instances of gains far exceeded declines. For the percentage proficient measure, there were 133 instances of moderate-to-large gains (65% of instances), compared with only 10 moderate-to-large declines (5% of instances). The effect size indicator had a similar result.

Table 3-A contains a total of 374 instances. This is fewer than the maximum number of possible instances because some states made changes to their testing programs at one or more grades and did not have three years of trend data. In addition, some states could provide percentage proficient data but not the mean test scores or standard deviations necessary to compute effect sizes; trends in these states were counted on one instead of two indicators.

Table 3-A. Instances of Gains and Declines in State Reading and Mathematics Test Scores by Indicator Since 2002

	Number (*percentage*) of instances of					Total number of instances
	↑ Moderate-to-large gains	↗ Slight gains	✪ No change	↘ Slight declines	↓ Moderate-to-large declines	
Percentage proficient	133 *(65%)*	43 *(21%)*	7 *(3%)*	13 *(6%)*	9 *(4%)*	205
Effect size	121 *(72%)*	10 *(6%)*	18 *(11%)*	9 *(15%)*	11 *(7%)*	169

Table reads: Of all the achievement trends analyzed in all states at three grade levels, in two subjects, and on two indicators, there were 133 instances of moderate-to-large gains and 9 instances of moderate-to-large declines in percentages proficient.

States with Consistent Trends Across Subjects and Grade Levels

Table 3-B focuses on a specific set of states—those that have sufficient comparable data for both the percentage proficient and effect size indicators and have made consistent gains across grade levels since NCLB took effect in 2002. The table reveals several interesting trends:

- Four states (Arkansas, Kentucky, Texas, and Washington) showed consistent moderate-to-large gains in test scores since 2002, out of the 25 states with sufficient data to discern trends in both reading and math. The gains were found across the elementary, middle, and high school levels in both subjects using both the percentage proficient and effect size indicators. In other words, achievement in these four states has improved across the board. Several other states posted consistent gains in one subject area but not the other.

- In several states, the gains varied in magnitude, with some slight and some moderate-to-large gains across grade levels, subject areas, and indicators. For example, a state might have had a moderate-to-large gain at a certain grade level using the percentage proficient indicator but only a slight gain using the effect size indicator. No patterns emerged of these states doing better on one indicator or the other.

- None of these states showed a decline of any magnitude across all three grade spans in either reading or math.

Table 3-B. Test Score Trends Since 2002 Across Three Grade Levels
(Includes only those states with both percentage proficient and effect size data)

	Type of Trend	All three grade levels		
		Both Reading & Math	Reading	Math
Both percentage proficient and effect size	↑ Moderate-to-large gains	4 states: AR, KY, TX, WA	6 states: AR, KY, ND, TN, TX, WA	9 states: AR, KY, LA, MS, NJ, OH, OK, TX, WA
	↑/↗ Mixed moderate-to-large and slight gains	1 state: UT	3 states: LA, ID, UT	5 states: IN, ND, PA, UT, WV
	↗ Slight gains	0 states	0 states	0 states
	Number of states with sufficient trend data for this analysis	25 states: AZ, AR, CA, CO, FL, HI, ID, IN, IA, KY, LA, MS, NV, NJ, ND, OH, OK, OR, PA, SC, TN, TX, UT, WA, WV	26 states: AZ, AR, CA, CO, FL, HI, ID, IN, IA, KY, LA, MS, NV, NJ, ND, OH, OK, OR, PA, SC, SD, TN, TX, UT, WA, WV	25 states: AZ, AR, CA, CO, FL, HI, ID, IN, IA, KY, LA, MS, NV, NJ, ND, OH, OK, OR, PA, SC, TN, TX, UT, WA, WV

Table reads: Since NCLB was enacted in 2002, four states made moderate-to-large gains in reading and math at all three grade levels analyzed (elementary, middle, and high school) according to both the percentages of students scoring at or above the proficient level and effect sizes. Six states have made moderate-to-large gains at all three grade levels in reading, and nine have made moderate-to-large gains at all three grade levels in mathematics.

Results by Grade Level in Reading

In addition to looking across grade levels, we analyzed gains and declines in state reading test scores since 2002 at each of the three grade levels—elementary, middle, and high school. The tables in this section present trends in states that have data on percentages proficient, effect sizes, or both. Also shown are states that did not have sufficient data for either indicator. The number of states with sufficient trend data for percentages proficient and/or effect sizes in reading totaled 35 at the elementary level, 35 at the middle school level, and 37 at the high school level.

Box 3-A. A Note on State Enrollments

Many of the results in this report are presented in terms of the numbers of states showing gains and declines in student achievement as measured by state tests and NAEP. However, users of the information in this report should also consider the relative number of students in each state. The table below ranks states by enrollments.

Trends in some states affect many more students than trends in other states. If large states show trends that are different from those in the rest of the country, then a different picture of overall student achievement in the nation might emerge. To address this concern, we examined achievement trends in the 20 states with the highest enrollments, which account for 75% of the student population nationally. We found that the trends for those states were not notably different from the trends we found for the rest of the country: gains in test scores far outnumbered declines, and gains tended to be smaller at the high school level than at the lower grades. In addition, the proportion of states with sufficient data for our analyses was not different in the 20 highest-enrollment states than it was in the 30 lower-enrollment states.

We also looked more closely at the five states with the largest K–12 enrollments (California, Texas, New York, Florida, and Illinois). Together, these states account for just over a third of U.S. student enrollment. Two states (New York and Illinois) did not have three years of trend data. California, Texas, and Florida had results which, taken together, reflect the national picture: an upward trend for most grade levels with a few declines, and elementary and middle school levels doing better than the high school level. (Trends for these and other states are shown in table 3-I, later in this chapter.) Overall, what happened in the large states reflected the results for all states.

Student Enrollment by State, 2006-07

Ranking	State	Enrollment
1	CA	6,286,943
2	TX	4,576,933
3	NY	2,741,258
4	FL	2,605,976
5	IL	2,043,001
6	PA	1,821,383
7	OH	1,803,226
8	MI	1,675,234
9	GA	1,589,839
10	NC	1,434,625
11	NJ	1,387,963
12	VA	1,221,939
13	IN	1,154,797
14	AZ	1,120,520
15	WA	1,028,377
16	MA	968,661
17	TN	925,898
18	MO	899,558
19	MD	851,640
20	WI	843,879
21	MN	828,243
22	CO	802,639
23	AL	739,760
24	SC	681,790
25	KY	668,337
26	LA	652,394
27	OK	633,006
28	UT	613,210
29	OR	562,828
30	CT	561,091
31	MS	489,071
32	IA	474,867
33	AR	465,615
34	KS	465,105
35	NV	425,731
36	NM	325,731
37	NE	287,135
38	WV	281,298
39	ID	276,723
40	NH	200,996
41	ME	187,450
42	HI	179,234
43	RI	149,855
44	MT	144,418
45	AK	130,776
46	DE	122,261
47	SD	121,090
48	VT	95,481
49	ND	94,057
50	WY	84,611

Table reads: Pennsylvania is the sixth largest state in terms of K–12 student enrollment, with 1,821,383 students.

Source: Center on Education Policy, based on data collected from state departments of education, winter 2007-08.

ELEMENTARY SCHOOL READING

As displayed in **table 3-C**, the majority of states with sufficient data showed moderate-to-large gains in elementary school reading, whether states had percentages proficient only, effect sizes only, or both. Of the 28 states with data for both indicators, 17 posted moderate-to-large gains on both. Of the 6 states with percentage proficient data only, 5 demonstrated moderate-to-large gains. For some states, percentages proficient gave a more positive picture of reading achievement than effect sizes. No state showed a decline in percentages proficient. Three states (Arizona, Indiana, and New Mexico) exhibited declines in effect sizes, and in all 3 of these states percentages proficient increased.

Table 3-C. Test Score Trends Since 2002 in Elementary School Reading

		Effect size					
Percentage of students scoring at or above the proficient level	Type of Trend	↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	♦/○ Data unavailable or too few years
	↑ Moderate-to-large gain	↑ ↑ 17 states: AR, CA, FL, ID, IA, KY, LA, MS, NV, ND, OH, OK, SC, TN, TX, WA, WV		↑ ✪ 1 state: OR	↑ ↘ 1 state: NM		↑ ♦/○ 5 states: AL, AK, MD, MT, NE
	↗ Slight gain	↗ ↑ 2 states: NC, UT		↗ ✪ 5 states: CO, HI, NJ, PA, SD	↗ ↘ 1 state: IN	↗ ↓ 1 state: AZ	↗ ♦/○ 1 state: MA
	✪ No change						
	↘ Slight decline						
	↓ Moderate-to-large decline						
	♦/○ Data unavailable or too few years	♦/○ ↑ 1 state: DE					♦/○ ♦/○ 15 states: CT, GA, IL, KS, ME, MI, MN, MO, NH, NY, RI, VT, VA, WI, WY

Table reads: Since NCLB was enacted in 2002, 17 states have made moderate-to-large gains in reading at the elementary grade analyzed, according to both the percentages of students scoring proficient and effect sizes.

MIDDLE SCHOOL READING

Of the 28 states with both percentage proficient and effect size data in reading, 14 states demonstrated moderate-to-large gains in middle school reading on both indicators (see **table 3-D**). All 6 states with percentage proficient data only made moderate-to-large gains. There were more declines in reading scores at the middle school level than at the elementary level, but there were still far fewer declines than gains. Percentages proficient decreased in 4 states and effect sizes decreased in 4 states; in 2 of these states, test scores fell on both indicators.

Table 3-D. Test Score Trends Since 2002 in Middle School Reading

		Effect size					
	Type of Trend	↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	♦/○ Data unavailable or too few years
Percentage of students scoring at or above the proficient level	↑ Moderate-to-large gain	↑ ↑ 14 states: AR, CA, ID, IN, IA, KY, LA, NV, NM, ND, PA, TN, TX, WA					↑ ♦/○ 6 states: AL, AK, MD, MA, MT, NE
	↗ Slight gain	↗ ↑ 5 states: CO, FL, MS, OR, UT	↗ ↗ 1 state: NC	↗ ✪ 1 state: OK	↗ ↘ 1 state: OH	↗ ↓ 1 state: NJ[1]	
	✪ No change			✪ ✪ 1 state: WV			
	↘ Slight decline			↘ ✪ 1 state: SC	↘ ↘ 1 state: AZ	↘ ↓ 1 state: HI	
	↓ Moderate-to-large decline			↓ ✪ 1 state: SD			
	♦/○ Data unavailable or too few years	♦/○ ↑ 1 state: DE					♦/○ ♦/○ 15 states: CT, GA, IL, KS, ME, MI, MN, MO, NH, NY, RI, VT, VA, WI, WY

Table reads: Since NCLB was enacted in 2002, 14 states have made moderate-to-large gains in reading at the middle school grade analyzed, according to both the percentages of students scoring proficient and effect sizes.

[1] The percentage proficient data covered a different range of years than the effect size data, which my explain the contradictory finding.

HIGH SCHOOL READING

Of the 27 states with both percentages proficient and effect sizes, only 8 states made moderate-to-large gains in high school reading on both indicators (see **table 3-E**). This is fewer than the number of states with this degree of gain at the elementary and middle school levels. Of the 9 states with percentage proficient data only, 5 demonstrated moderate-to-large reading gains at the high school level. Altogether, there were fewer gains and more declines in reading at the high school than at the elementary and middle school levels. Achievement declined on both percentages proficient and effect sizes at a moderate-to-large rate in 2 states and at a slight rate in 3 states. Several other states showed declines on one of the two indicators. The percentage proficient and the effect size indicators generally showed the same trends. But in 3 states (Ohio, Hawaii, and Nevada) trends went up on one indicator and down on the other.

Table 3-E. Test Score Trends Since 2002 in High School Reading

		Effect size					
Percentage of students scoring at or above the proficient level	Type of Trend	↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	♦/○ Data unavailable or too few years
	↑ Moderate-to-large gain	↑ ↑ 8 states: AR, KY, ND, OK, PA, TN, TX, WA				↑ ↓ 1 state: OH	↑ ♦/○ 5 states: MD, MA, MT, NE, NH
	↗ Slight gain	↗ ↑ 6 states: CO, ID, NJ, OR, SC, UT	↗ ↗ 1 state: LA	↗ ✪ 1 state: CT	↗ ↘ 1 state: HI		↗ ♦/○ 1 state: RI
	✪ No change			✪ ✪ 1 state: IN			
	↘ Slight decline			↘ ✪ 2 states: CA, IA	↘ ↘ 3 states: AZ, FL, WV		↘ ♦/○ 1 state: AL
	↓ Moderate-to-large decline		↓ ↗ 1 state: NV			↓ ↓ 2 states: MS, SD	↓ ♦/○ 2 states: ME, NM
	♦/○ Data unavailable or too few years	♦/○ ↑ 1 state: DE					♦/○ ♦/○ 13 states: AK, GA, IL, KS, MI, MN, MO, NY, NC, VT, VA, WI, WY

Table reads: Since NCLB was enacted in 2002, eight states have made moderate-to-large gains in reading at the high school grade analyzed, according to both the percentages of students scoring proficient and effect sizes.

Results by Grade Level in Mathematics

The number of states with sufficient data in math on either or both indicators totaled 34 at the elementary level, 34 at the middle school level, and 36 at the high school level.

ELEMENTARY SCHOOL MATH

Overwhelmingly positive trends were apparent in elementary school mathematics, as **table 3-F** illustrates. In 21 of the 27 states with both percentages proficient and effect sizes, achievement increased at a moderate-to-large rate. Five of the 6 remaining states with both types of data showed moderate-to-large gains in effect sizes and slight gains in percentages proficient. All 7 states with just one type of data had moderate-to-large gains.

Table 3-F. Test Score Trends Since 2002 in Elementary School Math

		Effect size					
	Type of Trend	↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	♦/❍ Data unavailable or too few years
Percentage of students scoring at or above the proficient level	↑ Moderate-to-large gain	↑ ↑ 21 states: AZ, AR, CA, FL, GA, ID, IA, KY, LA, MS, NV, NJ, NM, OH, OK, PA, SC, TN, TX, WA, WV				↑ ↓ 1 state: OR[1]	↑ ♦/❍ 6 states: AL, AK, MD, MA, MT, NE
	↗ Slight gain	↗ ↑ 5 states: CO, HI, IN, ND, UT					
	✪ No change						
	↘ Slight decline						
	↓ Moderate-to-large decline						
	♦/❍ Data unavailable or too few years	♦/❍ ↑ 1 state: DE					♦/❍ ♦/❍ 16 states: CT, IL, KS, ME, MI, MN, MO, NH, NY, NC, RI, SD, VT, VA, WI, WY

Table reads: Since NCLB was enacted in 2002, 21 states have made moderate-to-large gains in mathematics at the elementary grade analyzed, according to both the percentages of students scoring proficient and effect sizes.

[1] The percentage proficient data covered a different range of years than the effect size data, which may explain the contradictory finding.

MIDDLE SCHOOL MATH

We also found positive trends in middle school mathematics, as **table 3-G** reveals. Of the 27 states with sufficient data, 22 made moderate-to-large gains on both indicators. Of the 7 states with only one type of data, 6 demonstrated moderate-to-large increases.

Table 3-G. Test Score Trends Since 2002 in Middle School Math

		Effect size					
Percentage of students scoring at or above the proficient level	Type of Trend	↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	♦/○ Data unavailable or too few years
	↑ Moderate-to-large gain	↑ ↑ 22 states: AR, CO, FL, GA, HI, ID, IN, IA, KY, LA, MS, NJ, NM, OH, OK, OR, PA, TN, TX, UT, WA, WV	↑ ↗ 1 state: NY				↑ ♦/○ 5 states: AL, AK, MD, MA, NE
	↗ Slight gain		↗ ↗ 2 states: ND, SC				
	✪ No change				✪ ↘ 1 state: AZ		
	↘ Slight decline	↘ ↑ 1 state: CA[1]					
	↓ Moderate-to-large decline						↓ ♦/○ 1 state: MT
	♦/○ Data unavailable or too few years	♦/○ ↑ 1 state: DE					♦/○ ♦/○ 16 states: CT, IL, KS, ME, MI, MN, MO, NH, NY, NC, RI, SD, VT, VA, WI, WY

Table reads: Since NCLB was enacted in 2002, 22 states have demonstrated moderate-to-large gains in mathematics at the middle school grade analyzed, according to both the percentages of students scoring proficient and effect sizes. Two states showed slight gains on both indicators.

[1] The percentage proficient data covered a different range of years than the effect size data, which may explain the contradictory finding.

HIGH SCHOOL MATH

The high school math results summarized in **table 3-H** showed more declines than in the earlier grades, although gains still outnumbered declines. Of the 26 states with sufficient data for both percentages proficient and effect sizes, 12 states evidenced moderate-to-large gains on both indicators in high school math—far fewer than the number with math gains in elementary and middle school. In 6 states, achievement declined to some degree on one or both indicators. In Nevada, one of 2 states with contradictory trends on the two indicators, the percentage proficient decreased while effect sizes increased in math. In Idaho, the other such state, the percentage proficient in high school math rose slightly while effect size declined (although data was insufficient in Idaho to compute the 2007 effect size).

Table 3-H. Test Score Trends Since 2002 in High School Math

		Effect size						
	Type of Trend	↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	♦/❍ Data unavailable or too few years
Percentage of students scoring at or above the proficient level	↑ Moderate-to-large gain	↑ ↑ 12 states: AR, CA, KY, LA, MS, NJ, ND, OH, OK, TX, UT, WA	↑ ↗ 1 state: WV	↑ ✪ 1 state: FL			↑ ♦/❍ 5 states: AL, ME, MA, NE, NH
	↗ Slight gain	↗ ↑ 1 state: IN	↗ ↗ 1 state: PA	↗ ✪ 2 states: CT, SC		↗ ↓ 1 state: ID[1]	↗ ♦/❍ 2 states: GA, NM
	✪ No change		✪ ↗ 2 states: CO, IA			✪ ↓ 2 states: OR, TN	
	↘ Slight decline			↘ ✪ 1 state: HI		↘ ↓ 1 state: AZ	↘ ♦/❍ 1 state: RI
	↓ Moderate-to-large decline	↓ ↑ 1 state: NV					↓ ♦/❍ 1 state: MT
	♦/❍ Data unavailable or too few years	♦/❍ ↑ 1 state: DE					♦/❍ ♦/❍ 14 states: AK, IL, KS, MD, MI, MN, MO, NY, NC, SD, VT, VA, WI, WY

Table reads: Since NCLB was enacted in 2002, 12 states have made moderate-to-large gains in mathematics at the high school grade analyzed, according to both the percentages of students scoring proficient and effect sizes.

[1] The percentage proficient data covered a different range of years than the effect size data, which may explain the contradictory finding.

State-by-State Summary of Overall Achievement Trends

Table 3-I summarizes our findings about overall achievement for each state, displaying in more detail the specific state trends on which the tallies in tables 3-A through 3-H of this chapter are based. This table is not intended to support comparisons between states. Each state has its own standards, tests, proficiency definitions, and cut scores. Because of differences in testing systems, some states may be more likely to show gains than others.

Table 3-I. State-by-State Summary of Overall Achievement Trends

Legend

PP = Percentage proficient

ES = Effect size

↑ = Moderate-to-large gain (average yearly gain of ≥ 1.0 percentage point or ≥ 0.02 ES)

↓ = Moderate-to-large decline (average yearly decline of ≥ 1.0 percentage point or ≥ 0.02 ES)

↗ = Slight gain (average yearly gain of < 1 percentage point or < 0.02 ES)

↘ = Slight decline (average yearly decline of < 1 percentage point or < 0.02 ES)

✪ = No change (average yearly change of 0.0 percentage point or 0.00 ES)

♦ = Not enough years of data (only 1-2 years) to determine trend

❍ = Data not available

State & Years Analyzed	Reading						Mathematics					
	Elementary		Middle		High		Elementary		Middle		High	
	PP	ES	PP	ES	PP	ES	PP	ES	PP	ES	PP	ES
Alabama *(2004–2007)*	↑	♦	↑	♦	↘	♦	↑	♦	↑	♦	↑	♦
Alaska *(2005–2007)*	↑	♦	↑	♦	♦	♦	↑	♦	↑	♦	♦	♦
Arizona *(2005–2007)*	↗	↓	↘	↘	↘	↘	↑	↑	✪	↘	↘	↓
Arkansas *(2002–2007)*	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
California *(2003–2007)*	↑	↑	↑	↑	↘	✪	↑	↑	↘	↑	↑	↑
Colorado *(2002–2007)*	↗	✪	↗	↑	↗	↑	↗	↑	↑	↑	✪	↗
Connecticut *(2002–2006)*	♦	♦	♦	♦	↗	✪	♦	♦	♦	♦	↗	✪
Delaware *(2002–2007)*	♦	↑	♦	↑	♦	↑	♦	↑	♦	↑	♦	↑
Florida *(2002–2007)*	↑	↑	↗	↑	↘	↘	↑	↑	↑	↑	↑	✪
Georgia *(2002–2007)*	♦	♦	♦	♦	♦	❍	↑	↑	↑	↑	↗	❍
Hawaii *(2002–2006)*	↗	✪	↘	↓	↗	↘	↗	↑	↑	↑	↘	✪
Idaho *(2003–2007)*	↑	↑	↑	↑	↗	↑	↑	↑	↑	↑	↗	↓
Illinois *(2006–2007)*	♦	♦	♦	♦	♦	♦	♦	♦	♦	♦	♦	♦
Indiana *(2002–2007)*	↗	↘	↑	↑	✪	✪	↗	↑	↑	↑	↗	↑
Iowa *(2004–2007)*	↑	↑	↑	↑	↘	✪	↑	↑	↑	↑	✪	↗
Kansas *(2006–2007)*	♦	♦	♦	♦	♦	♦	♦	♦	♦	♦	♦	♦
Kentucky *(2002–2006)*	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
Louisiana *(2002–2007)*	↑	↑	↑	↑	↗	↗	↑	↑	↑	↑	↑	↑
Maine *(2003–2007)*	♦	♦	♦	♦	↓	❍	♦	♦	♦	♦	↑	❍

Continued on page 37

Table 3-I. *continued*

State & Years Analyzed	Reading						Mathematics					
	Elementary		Middle		High		Elementary		Middle		High	
	PP	ES	PP	ES	PP	ES	PP	ES	PP	ES	PP	ES
Maryland *(2003–2007)*	↑	○	↑	○	↑	○	↑	○	↑	○	◆	○
Massachusetts *(2002–2007)*	↗	○	↑	○	↑	○	↑	○	↑	○	↑	○
Michigan *(2006–2007)*	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆
Minnesota *(2006–2007)*	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆
Mississippi *(2002–2007)*	↑	↑	↗	↑	↓	↓	↑	↑	↑	↑	↑	↑
Missouri *(2006–2007)*	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆
Montana *(2004–2007)*	↑	◆	↑	◆	↑	◆	↑	◆	↓	◆	↓	◆
Nebraska *(2002–2007)*	↑	○	↑	○	↑	○	↑	○	↑	○	↑	○
Nevada *(2004–2007)*	↑	↑	↑	↑	↓	↗	↑	↑	↑	↗	↓	↑
New Hampshire *(2002–2007)*	◆	○	◆	○	↑	○	◆	○	◆	○	↑	○
New Jersey *(2002–2007)*	↗	✪	↗	↓	↗	↑	↑	↑	↑	↑	↑	↑
New Mexico *(2005–2007)*	↑	↘	↑	↑	↓	◆	↑	↑	↑	↑	↗	◆
New York *(2006–2007)*	◆	○	◆	○	○	○	◆	○	◆	○	○	○
North Carolina *(2003–2007)*	↗	↑	↗	↗	○	○	◆	◆	◆	◆	○	○
North Dakota *(2005–2007)*	↑	↑	↑	↑	↑	↑	↗	↑	↗	↗	↑	↑
Ohio *(2004–2007)*	↑	↑	↗	↘	↑	↓	↑	↑	↑	↑	↑	↑
Oklahoma *(2002–2007)*	↑	↑	↗	✪	↑	↑	↑	↑	↑	↑	↑	↑
Oregon *(2002–2007)*	↑	✪	↗	↑	↗	↑	↑	↓	↑	↑	✪	↓
Pennsylvania *(2002–2007)*	↗	✪	↑	↑	↑	↑	↑	↑	↑	↑	↗	↗
Rhode Island *(2004–2006)*	◆	○	◆	○	↗	○	◆	○	◆	○	↘	○
South Carolina *(2002–2007)*	↑	↑	↘	✪	↗	↑	↑	↑	↗	↗	↗	✪
South Dakota *(2005–2007)*	↗	✪	↓	✪	↓	↓	◆	◆	◆	◆	◆	◆
Tennessee *(2004–2007)*	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	✪	↓
Texas *(2005–2007)*	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
Utah *(2004–2007)*	↗	↑	↗	↑	↗	↑	↗	↑	↑	↑	↑	↑
Vermont *(2006–2007)*	◆	○	◆	○	◆	○	◆	○	◆	○	◆	○
Virginia *(2006–2007)*	◆	○	◆	○	◆	○	◆	○	◆	○	◆	○
Washington *(2002–2007)*	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
West Virginia *(2004–2007)*	↑	↑	✪	✪	↘	↘	↑	↑	↑	↑	↑	↗
Wisconsin *(2006–2007)*	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆
Wyoming *(2006–2007)*	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆	◆

Table reads: In Arizona, the percentage of students scoring at the proficient level or above on state reading tests increased slightly from 2005 to 2007 at the elementary grade analyzed for this report. When measured in terms of effect size, achievement in elementary-level reading declined at a moderate-to-large rate during that period.

Size of Gains and Declines

Our study also examined the size of gains and declines in percentages proficient and effect sizes. In both cases, we focused on average yearly gains, as described earlier in this chapter.

GAINS AND DECLINES IN READING

Tables 3-J and **3-K** display the magnitude of the gains and declines in achievement on state reading tests since 2002, in terms of percentages proficient and effect sizes. In table 3-J, the first three columns contain the median percentage proficient across all states at each grade level analyzed in 2002 or the beginning of each state's trend line, which may have been a few years after 2002. The median is a middle point; equal numbers of test scores lie above and below it. Also shown are the lowest percentage proficient in any state (the minimum) and the highest in any state (the maximum). This information gives readers an idea of where states were in terms of percentages proficient at the beginning of our time frame and the rate at which test scores have changed.

For example, at the high school level in reading, the median percentage proficient for all states in 2002 (or soon thereafter) was 66%, which means that half the states had percentages proficient below 66% and the other half had percentages proficient above 66%. There was a wide range in states' starting percentages proficient—Georgia had 96% of its high school students at or above the proficient level in reading at the start of its trend line in 2006, whereas Kentucky began with 29% proficient in 2002. (This is probably more of an indication of the relative difficulty of these two states' proficiency definitions than the achievement of their students. The full online achievement profiles for both states are available at www.cep-dc.org.)

The middle three columns give the median average yearly gain in percentages proficient in the states with sufficient test data in reading at a particular grade level. For example, the average yearly gain at the high school level over the period covered by this study was 1.4 percentage points. To arrive at this figure, we averaged the yearly gains for each of the 24 states with high school trend data in reading. The result was an average yearly gain for each state with sufficient data, and 1.4 was the median of those figures (thus we have "median averages"). While 1.4 percentage points might sound like a small gain, over a period of 12 years this would accumulate to an increase of 16.8 percentage points, such as from 66% to about 83% proficient.

Table 3-K gives similar information for effect sizes in reading. At the elementary school level, the median of the average yearly gains in effect size was 0.06 in the 21 states with sufficient data. Again, while this may appear small, if this rate continues, test scores will go up by 0.72 of a standard deviation over 12 years. That is a sizable increase, equivalent to increasing the percentage of students scoring above the 2002 average from 50% in 2002 to 76% in 2014. Another way of characterizing the change is that a student who was at the median in 2014 would have a score equal to that of a student at the 76th percentile in 2002. Arkansas showed the maximum gain of about one-fifth of a standard deviation for middle school reading. The state had introduced a new test for grades 3-8 in 2005, and there was an unusually large jump in scores between the first and second years the test was administered.

Average yearly test score gains in reading were greater at the elementary and middle school levels than at the high school level. The median average yearly gain on both indicators was smaller at the high school level, and the median average yearly decline in percentages proficient was greatest at the high school level. Generally, the magnitude of gains was greater than that of declines.

Table 3-J. Statistics on Starting Percentage Proficient and Average Yearly Gains and Declines in Percentages Proficient in Reading

Statistic	Starting percentage proficient			Average Yearly Percentage Point Gain			Average Yearly Percentage Point Decline		
	Elem.	Middle	H.S.	Elem.	Middle	H.S.	Elem.	Middle	H.S.
Median	73%	65%	66%	1.7	1.9	1.4	NA	-0.6	-0.9
Minimum	34%	27%	29%	0.2	0.1	0.3	NA	-0.5	-0.1
Maximum	87%	89%	96%	5.0	7.0	8.5	NA	-1.0	-4.8
Number of states with sufficient data	50	50	50	34	29	24	0	4	11

Table reads: In reading, the median percentage proficient for the 50 states in the elementary school grade analyzed was 73% in 2002 (or a later year in some states, depending on the availability of trend data). The minimum percentage proficient in the state with the lowest starting percentage was 34%. For the 34 states with three or more years of comparable trend data in elementary reading, the median average yearly gain in the percentage proficient was 1.7 percentage points over the years analyzed.

Note: Average yearly gains are shown as positive numbers and average yearly declines as negative numbers.

Table 3-K. Statistics on Average Yearly Gains and Declines in Effect Size in Reading Since 2002

Statistic	Average Yearly Effect Size Gain			Average Yearly Effect Size Decline		
	Elem.	Middle	H.S.	Elem.	Middle	H.S.
Median	0.06	0.05	0.04	-0.01	-0.01	-0.01
Minimum	0.00	0.01	0.01	-0.01	-0.01	-0.01
Maximum	0.09	0.22	0.20	-0.02	-0.03	-0.09
Number of states with sufficient data	21	21	17	2	4	8

Table reads: Of the 21 states with sufficient data to determine trends, the median average yearly gain in effect size in reading was 0.06 standard deviations per year at the elementary level. Average yearly effect size gains ranged from a minimum of 0 to a maximum of 0.09 standard deviations.

Note: Average yearly gains are shown as positive numbers and average yearly declines as negative numbers.

MATHEMATICS GAINS AND DECLINES

Table 3-L shows the size of gains and declines in percentages proficient in mathematics, and **table 3-M** shows the same information for effect sizes. As in reading, results were not as positive at the high school level as they were at the lower levels. The median average yearly gain for both indicators in math was smaller at the high school level than at the two lower grade levels. The median average yearly percentage point decline was greatest at the middle school level, while the average yearly effect size decline was greatest at the high school level.

The maximum average yearly effect size gain in the state with the greatest gain was 0.34. Again, that was an atypical result from Arkansas, for the reason explained above.

Table 3-L. Statistics on Starting Percentage Proficient, Average Yearly Gains and Declines in Percentages Proficient in Mathematics

Statistic	Starting Percentage proficient			Average Yearly Percentage Point Gain			Average Yearly Percentage Point Decline		
	Elem.	Middle	H.S.	Elem.	Middle	H.S.	Elem.	Middle	H.S.
Median	**69%**	**59%**	**56%**	**2.6**	**2.1**	**1.3**	**NA**	**-0.8**	**-0.7**
Minimum	21%	19%	19%	0.3	0.1	0.1	NA	-0.3	-0.3
Maximum	90%	83%	91%	7.7	7.5	5.2	NA	-1.3	-2.0
Number of states with sufficient data	50	50	48	33	30	26	0	2	5

Table reads: In math, the median percentage proficient for the 50 states in the elementary grade analyzed was 69% in 2002 (or a later year in some states, depending on the availability of trend data). The minimum percentage proficient in the state with the lowest starting percentage was 21%. For the 33 states with three or more years of comparable trend data in elementary math, the median average yearly gain in the percentage proficient in elementary reading was 2.6 percentage points over the years analyzed.

Note: Average yearly gains are shown as positive numbers and average yearly declines as negative numbers.

Table 3-M. Statistics on Average Yearly Gains and Declines in Effect Size in Math Since 2002

Statistic	Average Yearly Effect Size Gain			Average Yearly Effect Size Decline		
	Elem.	Middle	H.S.	Elem.	Middle	H.S.
Median	**0.07**	**0.05**	**0.04**	**-0.03**	**-0.01**	**-0.04**
Minimum	0.03	0.01	0.01	-0.03	-0.01	-0.02
Maximum	0.23	0.34	0.14	-0.03	-0.01	-0.07
Number of states with sufficient data	28	28	20	1	1	4

Table reads: Of the 28 states with sufficient data to determine trends, the median average yearly gain in effect size in mathematics was 0.07 standard deviations per year at the elementary level. Average yearly effect size gains ranged from a minimum of 0.03 to a maximum of 0.23 standard deviations.

Note: Average yearly gains are shown as positive numbers and average yearly declines as negative numbers.

Possible Explanations

There are several possible explanations for the trends in overall achievement emerging from our analyses. Below we highlight the most likely explanations. The list is not exhaustive but, instead, represents the explanations most often mentioned in research dealing with test score trends. Any or all of these factors in combination may be contributing to the observed trends, and different explanations could apply to different states, districts, or schools.

GAINS IN READING AND MATHEMATICS PERFORMANCE

Taken together, we found gains in reading and math achievement in most states with sufficient data since NCLB took effect in 2002. The number of states with gains far outweighed the number with declines. Likely explanations for these increases in achievement include the following:

- **Increased learning.** The most hopeful and positive factor that may be driving the increase in test scores is that students are indeed learning more and are demonstrating this learning on state tests. We know from CEP's own research on NCLB, as well as other research on the effects of other accountability programs, that states and districts have made major efforts to improve student achievement. NCLB-related surveys of school district and state officials (CEP, 2005, 2006; Hamilton, et al., 2007) found that school systems have increased the use of test results to inform instruction, aligned curriculum and instruction with standards, increased the rigor of their curricula, and provided extra attention and instruction to lower-performing students and subgroups such as students with disabilities and English language learners. These reforms have been instituted not only in response to NCLB; states and districts have their own accountability systems or other reforms that precede or exist alongside the federal law.

- **More time and emphasis on reading and math.** According to CEP's studies of curriculum and instructional changes since the enactment of NCLB, 58% of school districts have increased instructional time for reading at the elementary level and 45% have increased instructional time for elementary school math (CEP, 2007b; 2008). These increases were often substantial, amounting to at least 75 extra minutes per week for these subjects and sometimes 150 minutes or more per week. These shifts in instructional time have come at a price: 44% of the districts we surveyed reported making substantial cuts in time for other subjects, including social studies, science, art and music, physical education, recess, or lunch.

- **More familiarity with tests.** Researchers have noted that when new tests are introduced, gains are more pronounced in the first few years, then level off (Linn, 2000; Stecher & Hamilton, 2002). Since the inception of NCLB a large number of states have made changes to their testing programs and have introduced new tests. Part of the reason for gains may simply be that educators and students have become familiar with the format and content of the tests. As educators become more familiar with the tests, they often adjust instruction in order to better prepare students. Teachers might familiarize students with test instructions, item formats, and the scoring criteria used to grade open-ended test questions. As long as these activities do not take up a disproportionate amount of instructional time, they can be appropriate. But when classroom instruction becomes too focused on narrow test preparation—at the expense of teaching the broader state content standards—it can lead to score inflation, discussed next.

- **Score inflation.** The concept of score inflation means that test scores can increase without real gains in student learning (Koretz, 2005). Under pressure to show strong results on tests, educators may spend an inordinate amount of instructional time drilling students on practice questions released from prior years' tests, focusing instruction on the limited subset of skills and knowledge that are most likely to show up on the test, and teaching test-taking tactics. When scores increase on a test, it indicates that students have learned to correctly answer the specific kinds of questions included on that particular test. But mastery of the tested material is also supposed to indicate that students have attained greater mastery of the broader domain (such as 4th grade reading or 8th grade mathematics) that the test is intended to represent. One way to check for score inflation is to see whether test score gains carry over to other tests covering the same subject area. In chapter 4, we examine the extent to which state test score gains are also reflected on the National Assessment of Educational Progress.

- **More lenient tests, scoring, or data analyses.** Some studies have shown that states have lowered standards or have made tests easier in response to the demands of NCLB (Cronin et al., 2007). In our study, we have tried to address this by adjusting for formal, publicly announced breaks in testing systems or changes in cut scores. But as explained in chapter 2, test results can still be subtly manipulated through a number of small decisions made by education officials that affect such factors as scoring, choosing test items and assigning them relative weights, and equating scores across different tests or test versions. These can amount to tinkering rather than making substantial or formal decisions. Faced with pressure to raise test scores, state education officials might err on the side of leniency. It is difficult to find evidence of these types of changes, however.

LESS POSITIVE TRENDS AT HIGH SCHOOL LEVEL

Our analyses of achievement trends in reading and math indicate that test scores have not increased in high school to the same extent they have in lower grades. Other studies have also noted that high school performance lags compared with performance in earlier grades (Education Trust, 2005). On the 2007 administration of NAEP, average scores for 17-year-olds have actually declined in science since the 1990s, and the 2005 administration of NAEP showed a downward trend in reading (National Center for Education Statistics, 2006; 2007c). Trends in high school math could not be determined because a new assessment was introduced in 2005. Over roughly the same period, students at grades 4 and 8 posted gains on NAEP in most subject areas. U.S. high school students also compare slightly unfavorably with students in other countries on international assessments. According to the 2006 Program on International Student Assessment (PISA), U.S. 15-year-olds score below the average for market-oriented democracies in tests of math and science (NCES, 2007a). U.S. 4th and 8th graders tend to do a little better in international comparisons (NCES, 2004).

While we found more states with gains in achievement than declines at the high school level, substantially more states showed gains at the lower grade levels. There are several possible reasons for the less positive trends at the high school level:

- **Less learning in high school.** The most pessimistic conclusion is that somehow instruction has not improved at the high school level as much as it has at the lower grades. Or it may be harder to make improvements in the achievement of high school students than for younger students. Psychological research has shown that students' intrinsic desire to learn becomes weaker with each advancing grade (Lepper & Hederlong, 2000). Other factors, such as school safety, may negatively affect how much students are learning in

high school. Although high schools nationwide are relatively safe, researchers have found higher levels of violence at high school than at lower grades (Anderson et al., 2001; Surgeon General, 2001).

- **Less motivation to perform well on tests.** Research has noted the lack of motivation for high school students to perform well on NAEP and international exams (Brophy & Ames, 2005). NAEP has no stakes for individual students, so perhaps they have no reason to try hard. But state exams often do have stakes for individual students. About half of the states have high school exit exams, which students must pass to receive a high school diploma, and many states also use these exams for NCLB accountability purposes. So lack of motivation is not so easy to understand in these cases.

- **Inadequate preparation at lower grades.** It may be harder for high school students to show gains because their instruction at earlier grades may not have prepared them adequately for success in high school. For example, high school students may not be performing well in algebra because of lack of preparation at the lower grades, a situation that Cronin et al. (2007) call a lack of "calibration" in curricula.

- **Course selection.** Students in high school have greater choice in the courses they take than students at the lower grades, and high school curriculum varies more across districts. As a result, merely because of the expanded course options, students may be tested in math when they are not taking a math course at the time. Or, they may be taking a math course that does not mesh with the content focus of the state math test. For example, many 10th graders may be taking a geometry course, but the state test may have only a small number of geometry items.

- **More Title I funds go to elementary schools than high schools.** About 74% of federal Title I funds go to elementary schools, and 72% of students receiving Title I services in 2004-05 were in pre-kindergarten though grade 6 (Stulich et al., 2007). These Title I funds are used to provide extra services for students not performing at grade level, and these extra resources may help explain the greater gains found at the elementary level (Education Trust, 2005).

CHAPTER 4

Comparing State Trends in Overall Achievement with NAEP Trends

Summary of Findings

When viewed in conjunction with results from state tests, results from the National Assessment of Educational Progress offer an independent perspective on whether students have made general improvements in learning that show up on different assessments. Our study compared the overall achievement trends on state tests discussed in the previous chapter with trends on NAEP between 2003 and 2007. For reasons explained later in this chapter, we compared trends in the percentage of students scoring at or above the proficient level on state tests with trends in the percentage scoring at or above the basic level on NAEP. We also compared trends in effect sizes on the two assessments. Here are our main findings from these comparisons:

- **Trends in reading and math achievement on NAEP have tended to move in the same direction as trends on state tests in the 29 or more states with sufficient data for these comparisons.** When average yearly changes in percentages proficient on state tests were compared with average yearly changes in percentages basic on NAEP, trends on the two assessments moved in the same direction 82% of the time. When average yearly changes in effect sizes were compared on state tests and NAEP, effect size trends followed the same direction 67% of the time.

- **The extent of agreement between NAEP trends and state test score trends differed by subject and grade level, with the greatest divergence of trends in grade 8 reading.** Trends on NAEP and state tests moved in the same direction more often in math than in reading—91% agreement in math versus 59% in reading. In addition, trends on the two assessments corresponded more often at the elementary level than at the middle school level, and more often on the percentage proficient/basic indicator than on the effect size indicator. A major contributor to these differences was the divergence between state tests and NAEP in grade 8 reading, where percentages proficient/basic trends agreed 62% of the time, and effect size trends agreed a mere 20% of the time.

- **NAEP assessments, like state tests, revealed more gains in achievement than declines.** We made 252 different comparisons between NAEP and state tests (two subjects, times two grade levels, times two indicators, times the number of states with sufficient data). Of the 134 comparisons made between state percentages proficient and NAEP percentages basic, gains were found on both assessments in 108 cases and declines on both in just 2 cases. In the remaining 24 cases, trends on the two assessments diverged. Of the 118 comparisons made between effect size trends on state tests and NAEP, gains occurred on both assessments in 77 cases, no net change on both in 2 cases, and declines on both in no case. In the remaining 39 cases, trends diverged.

- **Gains on NAEP tended to be smaller than gains on state tests.** This pattern of smaller gains on NAEP emerged in both reading and math across all grade levels analyzed, and for both the percentage proficient/basic and effect size comparisons. Of the 252 points of comparison in states with sufficient data on both assessments, state tests showed larger gains in 188 cases (75%), while NAEP showed larger gains in 55 cases (22%).

How We Compared Trends on State Tests and NAEP

The National Assessment of Educational Progress is the only national assessment program designed to measure what students know and can do in reading, mathematics, and other core academic subjects. Administered by the U.S. Department of Education, NAEP not only provides a national picture of U.S. student achievement over time, but also reports results by state and by subgroup.[3]

DESIGN DIFFERENCES BETWEEN STATE TESTS AND NAEP

Although NAEP offers an important perspective on achievement that is independent from state tests, it should not be treated as a "gold standard" that overrides or invalidates state test results. NAEP differs from state tests—to varying degrees, depending on the state—in the content assessed, the test format and technical specifications, the rigor of the achievement levels, and other fundamental features. While state tests are designed to measure how well students have learned a particular state's academic content standards, NAEP is not deliberately aligned to any state's standards and therefore may not assess what students are actually taught. Furthermore, NAEP, unlike state tests, is designed so that samples of students in each state (rather than all students) take only a portion of the larger assessment (rather than the entire test). Consequently, NAEP cannot produce scores for individual students or schools. Finally, NAEP results are not tied to specific consequences for individual students, teachers, schools, or districts, as state test scores are, so students may not be as motivated to perform their best on NAEP.

Despite these design differences, comparisons of trends on state tests and NAEP are informative because NAEP is the best source of independent national achievement data at the grade levels we studied. Moreover, trends on NAEP are widely reported, and it would leave a large set of unanswered questions if our study discussed state test results without considering NAEP.

APPROACH USED TO COMPARE STATE TEST AND NAEP RESULTS

We used the following approach to compare state test results with NAEP results:

- **Subjects and grades.** For state tests, we looked at results in reading and math for the same elementary and middle school grades analyzed in other chapters of this report, chosen according to consistent criteria explained in chapter 2. In most cases, these were grades 4 and 8, but in some states a different adjacent grade, such as grade 5 or 7, was analyzed because sufficient comparable data were not available at grade 4 or 8. For NAEP, we used results from the reading and math tests at grades 4 and 8. (Grade 12 NAEP results could not be included because they are not broken down by state.) We compared results from the chosen state elementary grade with those from grade 4 NAEP and from the state middle school grade with grade 8 NAEP.

[3] The results discussed in this chapter and chapter 6 are from the main NAEP assessment. NAEP also administers a separate long-term trend assessment that has been in place since the 1970s but is administered less often and is less relevant to the purpose of this study.

- **Years analyzed and number of states included.** For each state's test, we used the same span of years employed for the other analyses in this report. These spans varied by state, depending on which years between 2002 and 2007 had comparable test data. As with the other analyses in this report, the trend data reported in this chapter include only those states with three or more years of comparable state test data. The number of states with sufficient data differed, depending on the grade level, subject, and indicator (percentages proficient or effect sizes) being analyzed. For NAEP, we looked at changes in test results between the 2003 and 2007 NAEP administrations, the longest post-NCLB span of years for which data were available for all 50 states.

- **Comparisons of state percentages proficient with NAEP percentages basic.** Both state tests and NAEP report their results in terms of various achievement levels—such as basic, proficient, and advanced—but the definitions and names for these levels vary among states and between state tests and NAEP. As explained in **box 4-A**, we looked at a range of evidence which suggested that the NAEP definition of proficient performance typically represents a more ambitious and aspirational concept of competent performance than the definitions of proficiency used by most states (or by many other countries). Based on this evidence, we concluded it was most appropriate to compare percentages of students scoring at or above the proficient level on state tests with percentages scoring at or above the basic level on NAEP, rather than with NAEP percentages proficient.

- **Effect size comparisons.** In addition to analyzing percentages proficient/basic, we also compared average yearly changes in effect sizes on state tests with those on NAEP. Unlike percentages proficient, effect sizes do not depend on where cut scores are set. Effect sizes also pick up improvements among students below the proficient level, which is where many of the gains on NAEP have occurred, as well as improvements above the proficient level.

- **Use of averages.** We compared average yearly changes in state test results with average yearly changes in NAEP results. To calculate these averages, we divided the overall change in the percentage proficient (percentage basic, in the case of NAEP) or the overall change in effect size by the number of years covered by the trend. This approach ensured that even though the state test score trends and NAEP trends sometimes covered different spans of years, we were looking at the average rate of change reflecting a year's worth of progress on both assessments.

- **Same definitions for moderate-to-large and slight changes.** We used the same definitions employed elsewhere in this report to characterize gains or declines as moderate-to-large or slight, and applied them consistently to results from both state tests and NAEP. (As explained in chapter 2, these definitions were developed for this study. Neither NAEP nor state testing programs distinguish between these two magnitudes of gains or declines.) According to these definitions, average changes of at least 1.0 percentage point per year or at least 0.02 effect size units per year were considered "moderate-to-large," while average yearly changes below these thresholds were considered "slight."

- **No tests of statistical significance for changes.** NAEP tests a sample of students in each state and computes statistical estimates of student performance to generalize results to the entire state's population. The NAEP program is understandably careful to report the degree of confidence that data users should have in these estimates, and highlights shifts in performance only when they are statistically significant. State tests, by contrast, are administered to all students in a particular grade, and changes in state test results do not have to reach any particular level of statistical significance to be counted under the requirements of No Child Left Behind. In this study, we treated trends on NAEP in

much the same way as we interpreted trends on state tests. We did not complicate or constrain comparisons by limiting NAEP data to statistically significant changes. However, we did check the extent to which the moderate-to-large gains and declines in NAEP achievement identified in this chapter were reported as statistically significant by NAEP, and we found this occurred for 82% of the gains and declines identified.

Applying all of the above criteria, we compared the directions of trends on state tests and NAEP in two subjects (reading and math), at two grades (4 and 8), on two indicators (percentages proficient/basic and effect sizes), in all states with sufficient data. We also compared the size of gains or declines in achievement on state tests and NAEP. The following sections of this chapter describe our findings from these analyses and include tables for each type of comparison. A final section considers possible reasons why state tests and NAEP may show different results.

Box 4-A. Why Compare the State Proficient Level with the NAEP Basic Level?

Both NAEP and state testing programs report results using multiple levels of student achievement. NAEP has defined three achievement levels—basic, proficient, and advanced—but does not attach consequences to any of them. States are required by No Child Left Behind to establish three achievement levels on their state tests—often called basic, proficient, and advanced but sometimes labeled differently. However, only the percentage of students reaching the proficient level carries weight for federal accountability. To determine which achievement level on NAEP should be compared with percentages proficient on state tests, we looked at the premises underlying these levels and evidence from various studies. We concluded that the percentage proficient on state tests was most appropriately compared with the percentage basic on NAEP.

Differences in State and NAEP Definitions of Achievement Levels

Although the labels for the achievement levels on NAEP and state tests are similar, the definitions are often quite different. The NAEP definitions are not linked to mastery of any state's specific academic content standards and do not take into account where students are achieving now. In a sense, the NAEP definition of proficient is aspirational, signaling where students should be and embodying the knowledge and skills that the NAEP governing board believes should be included in a well-designed curriculum for that subject area. To reach the NAEP proficient level, students must demonstrate "solid academic performance" and "competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter." To reach the NAEP basic level, students must demonstrate "partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade" (National Center for Education Statistics, n.d.).

State definitions of proficient, by contrast, are tied to the state's content standards and vary considerably. In addition, because state tests, unlike NAEP, are used for high-stakes accountability purposes, states are under pressure to set realistic definitions of proficiency that take into account students' current level of achievement. In most states, it could be said that students who do not meet the proficient standard are being left behind in learning the state's current standards. This difference between the aspirational goals of NAEP and the current goals of states helps to explain why U.S. Secretary of Education Margaret Spellings remarked that the proficient level on state tests is more appropriately compared to the NAEP basic than to the NAEP proficient level (Dillon, 2005.)

Evidence from the Mapping Study

A 2007 study by the National Center for Education Statistics, which administers NAEP, provided evidence that in most states, the cut scores for proficient performance on state tests were less ambitious than the NAEP proficient level and often were closer to—or sometimes below—the NAEP basic level. This study mapped states' proficient cut scores onto the NAEP scoring scales in reading and math at grades 4 and 8.

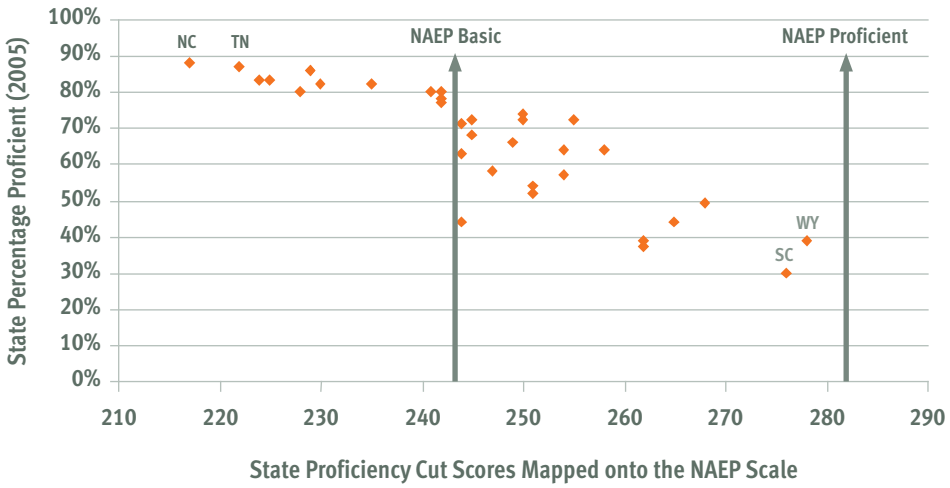
Continued on page 49

Box 4-A. (continued)

Figure 4-A shows the results of this mapping process for grade 8 reading. As the figure indicates, the state proficiency cut scores were below the NAEP basic level in many states and did not exceed the NAEP proficient level in any state. The figure also suggests that states with lower proficiency cut scores on state tests have higher percentages of students reaching the proficient level (displayed on the vertical axis), while states with more ambitious cut scores generally have lower percentages proficient. For instance, South Carolina has a relatively high cut score on its state grade 8 reading test (equivalent to a score of 275 on NAEP), and a relatively low percentage proficient (approximately 30%).

The NCES mapping study uncovered similar relationships for other grades and subjects, with state proficiency cut scores below the NAEP basic level in several states, between the NAEP basic and proficient levels in many states, and above the NAEP proficient level in a few or no states. These comparisons indicate that changes in the percentage of students reaching the proficient level on state tests correspond most closely to changes in the percentage reaching the basic level on NAEP.

Figure 4-A. Relationship of State Percentages Proficient to State Cut Scores for Proficient Performance, Mapped onto the NAEP Scoring Scale for Grade 8 Reading



Source: NCES, 2007.

Percentages of Students Reaching Various Performance Levels on State Tests and NAEP

To understand better the differences between achievement levels on NAEP and state tests, we also looked at the percentages of students scoring at or above the proficient level on their state test, the proficient level on NAEP, and the basic level on NAEP. **Table 4-A** gives a snapshot of these data in elementary and middle school reading from 2007, in most cases. For each of these categories, the table shows the median percentages of students reaching these various levels across all states. To give a sense of the range, the table also includes the lowest percentage in any state (the minimum) and the highest percentage in any state (the maximum).

Table 4-A provides additional evidence that proficient levels on state tests tend to be more comparable to the NAEP basic level. In elementary school reading, for example, the median percentage of students reaching their state's proficient level was higher than the median percentage reaching either the NAEP basic or NAEP proficient level. In elementary math, as well as in middle school reading and math, the median percentage of students reaching the state proficient level fell between the percentages meeting NAEP basic and NAEP proficient but was closer to the percentage meeting NAEP basic.

Continued on page 50

Box 4-A. (continued)

Table 4-A. Percentages of Students Scoring at or Above the Proficient Level on State Tests and at or Above the Basic and Proficient Levels on NAEP, 2007*

	Reading					
	Elementary/Grade 4			Middle/Grade 8		
	State Proficient	NAEP Basic	NAEP Proficient	State Proficient	NAEP Basic	NAEP Proficient
Median	79%	69%	34%	73%	76%	31%
Minimum	42%	51%	19%	25%	60%	17%
Maximum	90%	81%	49%	92%	85%	43%
	Mathematics					
	Elementary/Grade 4			Middle/Grade 8		
	State Proficient	NAEP Basic	NAEP Proficient	State Proficient	NAEP Basic	NAEP Proficient
Median	76%	84%	40%	66%	73%	34%
Minimum	24%	70%	21%	20%	54%	14%
Maximum	91%	93%	58%	88%	86%	51%

Table reads: Across states, the median percentage of students performing at or above the proficient level on their state's test was 79% in elementary school reading. On the NAEP grade 4 reading test, the median percentage of students performing at or above the NAEP basic level was 69% and the median performing at or above the NAEP proficient level was 34%.

* In the three states that initiated new tests or set new cut scores in 2007, these data were from 2006 instead of 2007.

Evidence from International Studies

The NAEP proficient level is also ambitious from an international perspective. One study by the American Institutes for Research linked the NAEP scoring scale to the scale of the Trends in International Mathematics and Science Study (TIMSS), an assessment of achievement in 46 countries including the U.S (Phillips, 2007). The study found that in only five countries, all Asian nations, did the average 8th grader reach the equivalent of the NAEP proficient level on TIMSS.

Direction of Achievement Trends

Our analyses sought to determine whether average yearly changes in achievement on NAEP between 2003 and 2007 had moved in the same direction as average yearly changes in achievement on state tests for the post-NCLB years with comparable data. We looked at both percentages proficient/basic and effect sizes from state tests and NAEP at two grade levels in reading and math.

LOOKING ACROSS GRADES AND SUBJECTS

Altogether, we made 252 separate comparisons between NAEP and state tests, including 134 in terms of percentages proficient/basic and 118 in terms of effect sizes. If all 50 states had had sufficient data on both assessments, we would have made eight comparisons per state—two subjects times two grade levels times two indicators—for a total of 400 comparisons. But the number of states with sufficient data for the specific comparisons ranged from 29 for effect sizes in math to 34 for percentages proficient/basic in reading, so the total was lower than 400.

Table 4-B shows how often overall achievement trends moved in the same direction on state tests and NAEP for states with sufficient data—in other words, how often test results went up on both assessments, showed no change on both, or declined on both. In essence, this table summarizes key data from the more detailed tables by subject, grade, and achievement indicator that follow.

As table 4-B indicates, there was a fair amount of consistency in the direction of trends on NAEP and state tests. This consistency was particularly striking for the percentages proficient/basic comparisons. Of the 134 comparisons using percentages proficient/basic, 108 comparisons showed slight or moderate-to-large gains on both state tests and NAEP, and just 2 showed slight or moderate-to-large declines on both assessments. (The remaining 24 comparisons revealed contradictory trends on NAEP and state tests.) In short, 82% of the percentage proficient/basic comparisons were in agreement about the direction of trends.

Comparisons using effect sizes were also consistent in direction in the majority of cases but with fewer cases of agreement than the percentage/proficient basic comparisons. Of the 118 comparisons using effect sizes, 77 revealed slight or moderate-to-large gains on both assessments, 2 showed no change on both, and none revealed declines on both, for an agreement rate of 67%. (The remaining 39 comparisons showed contradictory trends on the two assessments).

Using the data in table 4-B, one can also determine the extent of agreement on both assessments by subject and grade level. Trends agreed much more often in math than in reading—91% agreement versus 59%. Trends also agreed more often at grade 4 than at grade 8—86% agreement versus 64%.

A major contributor to these differences was the divergence between state tests and NAEP in grade 8 reading. Here, trends in percentages proficient/basic agreed 62% of the time, and trends in effect sizes agreed a mere 20% of the time.

Table 4-B. Agreement on Overall Achievement Trends Between State Tests and NAEP

Subject and Grade	Gains on Both	No Change on Both	Declines on Both	Total Instances of Agreement	Number of States with Data for Comparison	Percentage of Comparisons in Agreement
Percentages Proficient/Basic						
Grade 4 reading	30	0	0	30	34	88%
Grade 8 reading	19	0	2	21	34	62%
Grade 4 math	31	0	0	31	33	94%
Grade 8 math	28	0	0	28	33	85%
				110	**134**	**82%**
Effect Sizes						
Grade 4 reading	18	1	0	19	30	63%
Grade 8 reading	5	1	0	6	30	20%
Grade 4 math	28	0	0	28	29	97%
Grade 8 math	26	0	0	26	29	90%
				79	**118**	**67%**

Table reads: In elementary/grade 4 reading, 30 states demonstrated gains in both the percentage of students scoring proficient on state tests and the percentage scoring at the basic level on NAEP; none of the states with sufficient data had declines on both assessments or no change on both. Among the 34 states with sufficient data to compare state test and NAEP trends in grade 4 reading, trends on both assessments showed agreement in 30 states, or 88% of the time.

GRADE 4 READING, PERCENTAGES PROFICIENT/BASIC

At the elementary level, our analyses revealed strong correspondence between state and NAEP achievement trends and a general pattern of improving achievement. All of the 34 states with sufficient state test data in elementary school (usually grade 4) reading demonstrated gains on state tests since 2002, including 24 with moderate-to-large gains and 10 with slight gains. Of this group, 30 also showed gains on the NAEP grade 4 reading test, as depicted in **table 4-C**, including 19 with moderate-to-large gains on NAEP and 11 with slight gains. Only 4 of the states with gains on state tests demonstrated slight declines on NAEP, and none had moderate-to-large declines.

Table 4-C. Comparison of Percentage Proficient/Basic Trends on State Tests and NAEP in Elementary/Grade 4 Reading

		NAEP percentage basic					
	Type of Trend	↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	Total
State test percentage proficient	↑ Moderate-to-large gain	↑ ↑ 15 states: AK, AL, AR, FL, ID, KY, MD, MT, ND, NE, N.M, NV, O7H, OK, TX	↑ ↗ 6 states: CA, IA, LA, MS, TN, WA		↑ ↘ 3 states: OR, SC, WV		24
	↗ Slight gain	↗ ↑ 4 states: HI, MA, NJ, PA	↗ ↗ 5 states: AZ, CO, IN, SD, UT		↗ ↘ 1 state: NC		10
	✪ No change						0
	↘ Slight decline						0
	↓ Moderate-to-large decline						0
	Total	19	11	0	4	0	34

Table reads: Twenty-four states demonstrated moderate-to-large gains on state elementary school reading tests in the post-NCLB years with comparable data. Of these, 15 states also showed moderate-to-large gains on NAEP grade 4 reading tests between 2003 and 2007, while 6 states made slight gains on NAEP and 3 had slight declines on NAEP.

Note: The following 16 states were not included in this comparison with NAEP because data from state tests were unavailable or covered too few years to constitute a trend: CT, DE, GA, IL, KS, ME, MI, MN, MO, NH, NY, RI, VA, VT, WI, WY.

GRADE 4 READING, EFFECT SIZES

Average yearly trends in effect size on NAEP and state tests diverged more often than trends in percentages proficient/basic. Still, the agreement between the two types of assessments was fairly high in elementary/grade 4 reading. As illustrated in **table 4-D**, 21 of the 30 states with sufficient data showed gains on their elementary state reading tests since 2002; 18 of these states also posted gains on the NAEP grade 4 reading test between 2003 and 2007. On both state tests and NAEP, more states showed no net change or declines in achievement using effect sizes than using percentages proficient/basic (as displayed in table 4-C).

Table 4-D. Comparison of Effect Size Trends on State Tests and NAEP in Elementary/Grade 4 Reading

		NAEP effect size					
	Type of Trend	↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	Total
State test effect size	↑ Moderate-to-large gain	↑ ↑ 15 states: AR, CA, FL, ID, KS, KY, LA, MS, ND, NV, OH, OK, TN, TX, WA	↑ ↗ 3 states: DE, IA, UT		↑ ↘ 1 state: SC	↑ ↓ 2 states: NC, WV	21
	↗ Slight gain						0
	✪ No change	✪ ↑ 3 states: HI, NJ, PA	✪ ↗ 2 states: IN, SD	✪ ✪ 1 state: CO		✪ ↓ 1 state: OR	7
	↘ Slight decline	↘ ↑ 1 state: NM					1
	↓ Moderate-to-large decline			↓ ✪ 1 state: AZ			1
	Total	19	5	2	1	3	30

Table reads: Twenty-one states demonstrated moderate-to-large average annual gains in effect size on state elementary school reading tests in the post-NCLB years with comparable data. Of these, 15 states also showed moderate-to-large average annual gains in effect size on NAEP grade 4 reading tests between 2003 and 2007, while 3 states made slight gains on NAEP, 1 state had a slight decline on NAEP, and 2 states showed a moderate-to-large decline.

Note: The following 20 states were not included in this comparison with NAEP because data from state tests were unavailable or covered too few years to constitute a trend: AK, AL, CT, GA, IL, MA, MD, ME, MI, MN, MO, MT, NE, NH, NY, RI, VA, VT, WI, WY.

GRADE 8 READING, PERCENTAGES PROFICIENT/BASIC

The greatest discrepancies between state test score trends and NAEP trends were found in middle school (usually grade 8) reading. Thirty-four states had sufficient percentage proficient data to make comparisons for this grade and subject. As shown in **table 4-E**, 29 states made gains in percentages proficient on their state middle school reading tests; 19 of these states also made gains in percentages basic on the NAEP grade 8 reading test, although the majority of these NAEP gains were slight. The other 10 states with state test gains showed declines or no change on NAEP. Of the 4 states with declines on state tests, 2 posted slight gains on NAEP.

Table 4-E. Comparison of Percentage Proficient/Basic Trends on State Tests and NAEP in Middle School/Grade 8 Reading

		NAEP percentage basic					
	Type of Trend	↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	Total
State test percentage proficient	↑ Moderate-to-large gain	↑ ↑ 2 states: AK, MD	↑ ↗ 12 states: CA, IA, ID, LA, MA, MT, ND, NE, PA, TN, TX, WA	↑ ✪ 1 state: NM	↑ ↘ 4 states: AL, AR, IN, NV	↑ ↓ 1 state: KY	20
	↗ Slight gain	↗ ↑ 1 state: FL	↗ ↗ 4 states: CO, NJ, OH, OR		↗ ↘ 3 states: NC, OK, UT	↗ ↓ 1 state: MS	9
	✪ No change				✪ ↘ 1 state: WV		1
	↘ Slight decline		↘ ↗ 1 state: HI		↘ ↘ 2 states: AZ, SC		3
	↓ Moderate-to-large decline		↓ ↗ 1 state: SD				1
	Total	3	18	1	10	2	34

Table reads: Twenty states demonstrated moderate-to-large gains on state middle school reading tests in the post-NCLB years with comparable data. Of these, 2 states also showed moderate-to-large gains on NAEP grade 8 reading tests between 2003 and 2007, while 12 states made slight gains on NAEP, and 1 state had no net change on NAEP. Of this same group of 20 states, 4 states showed slight declines on NAEP, and 1 had a moderate-to-large decline.

Note: The following 16 states were not included in this comparison with NAEP because data from state tests were unavailable or covered too few years to constitute a trend: CT, DE, GA, IL, KS, ME, MI, MN, MO, NH, NY, RI, VA, VT, WI, WY.

GRADE 8 READING, EFFECT SIZES

In middle school reading, effect size comparisons also revealed little agreement between state test and NAEP results for the 30 states with sufficient data. As **table 4-F** demonstrates, 20 states posted moderate-to-large gains and 1 state posted slight gains on their state middle school reading tests, a generally positive picture. But effect size trends on the NAEP grade 8 reading test were less positive. Only 8 states made gains of any size on NAEP between 2003 and 2007. In fact, of the 20 states with moderate-to-large gains on state tests, 8 had no net change on NAEP, 4 had slight declines, and 3 had moderate-to-large declines.

Table 4-F. Comparison of Effect Size Trends on State Tests and NAEP in Middle School/Grade 8 Reading

		NAEP effect size					
	Type of Trend	↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	Total
State test effect size	↑ Moderate-to-large gain	↑ ↑ 3 states: FL, PA, TX	↑ ↗ 2 states: OR, TN	↑ ✪ 8 states: AR, CA, DE, IA, ID, LA, NV, WA	↑ ↘ 4 states: CO, IN, ND, NM	↑ ↓ 3 states: KY, MS, UT	20
	↗ Slight gain					↗ ↓ 1 state: NC	1
	✪ No change		✪ ↗ 1 state: KS	✪ ✪ 1 state: SD	✪ ↘ 1 state: SC	✪ ↓ 2 states: OK, WV	5
	↘ Slight decline		↘ ↗ 1 state: OH	↘ ✪ 1 state: AZ			2
	↓ Moderate-to-large decline	↓ ↑ 1 state: NJ		↓ ✪ 1 state: HI			2
	Total	4	4	11	5	6	30

Table reads: Twenty states demonstrated moderate-to-large average annual gains in effect size on state middle school reading tests in the post-NCLB years with comparable data. Of these, three states also showed moderate-to-large average annual gains in effect size on NAEP grade 8 reading tests between 2003 and 2007, while two states made slight gains on NAEP, eight showed no net change on NAEP, four had slight declines on NAEP, and three had moderate-to-large declines on NAEP.

Note: The following 20 states were not included in this comparison with NAEP because data from state tests were unavailable or covered too few years to constitute a trend: AK, AL, CT, GA, IL, MA, MD, ME, MI, MN, MO, MT, NE, NH, NY, RI, VA, VT, WI, WY.

GRADE 4 MATH, PERCENTAGES PROFICIENT/BASIC

The percentage proficient/basic analyses for elementary/grade 4 mathematics showed a clear pattern of gains on both state tests and NAEP and a strong correspondence between the two assessments. As shown in **table 4-G**, all 33 states with sufficient data made moderate-to-large or slight gains on their state elementary math tests during the years since 2002 with comparable data, and 31 of these states also demonstrated gains on the NAEP grade 4 math test between 2003 and 2007. Only 1 state showed a slight decline on NAEP, and one state reported no net change.

Table 4-G. Comparison of Percentage Proficient/Basic Trends on State Tests and NAEP in Elementary/Grade 4 Math

	Type of Trend	NAEP percentage basic: ↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	Total
State test percentage proficient	↑ Moderate-to-large gain	↑ ↑ 21 states: AK, AL, AR, FL, GA, ID, KY, LA, MA, MD, MS, MT, NJ, NM, NV, OH, OK, PA, TN, TX, WV	↑ ↗ 5 states: AZ, CA, IA, SC, WA	↑ ✪ 1 state: NE	↑ ↘ 1 state: OR		28
	↗ Slight gain	↗ ↑ 4 states: CO, HI, IN, ND	↗ ↗ 1 state: UT				5
	✪ No change						0
	↘ Slight decline						0
	↓ Moderate-to-large decline						0
	Total	25	6	1	1	0	33

Table reads: Twenty-eight states demonstrated moderate-to-large gains on state elementary math tests in the post-NCLB years with comparable data. Of these, 21 states also showed moderate-to-large gains on NAEP grade 4 math tests between 2003 and 2007, while 5 states made slight gains on NAEP, 1 state had no net change on NAEP, and 1 state showed a slight decline.

Note: The following 17 states were not included in this comparison with NAEP because data from state tests were unavailable or covered too few years to constitute a trend: CT, DE, IL, KS, ME, MI, MN, MO, NC, NH, NY, RI, SD, VA, VT, WI, WY.

GRADE 4 MATH, EFFECT SIZES

Effect size comparisons in elementary/grade 4 math paint a positive picture of gains and demonstrate the highest level of agreement between state test and NAEP results, as displayed in **table 4-H**. Of the 29 states with sufficient data for this comparison, 27 posted moderate-to-large gains on both their state test and the NAEP math test; 1 had a moderate-to-large gain on the state test and a slight gain on NAEP. The remaining state had a decline on the state test and no net change on NAEP.

Table 4-H. Comparison of Effect Size Trends on State Tests and NAEP in Elementary/Grade 4 Math

		NAEP effect size					
	Type of Trend	↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	Total
State test effect size	↑ Moderate-to-large gain	↑ ↑ 27 states: AR, AZ, CA, CO, DE, FL, GA, HI, IA, ID, IN, KS, KY, LA, MS, ND, NJ, NM, NV, OH, OK, PA, TN, TX, UT, WA, WV	↑ ↗ 1 state: SC				28
	↗ Slight gain						0
	✪ No change						0
	↘ Slight decline						0
	↓ Moderate-to-large decline			↓ ✪ 1 state: OR			1
	Total	27	1	1	0		29

Table reads: Twenty-eight states demonstrated moderate-to-large average annual gains in effect size on state elementary school math tests in the post-NCLB years with comparable data. Of these, 27 states also showed moderate-to-large average annual gains in effect size on NAEP grade 4 math tests between 2003 and 2007, while 1 state made a slight gain on NAEP. No state had a decline on NAEP.

Note: The following 21 states were not included in this comparison with NAEP because data from state tests were unavailable or covered too few years to constitute a trend: AK, AL, CT, IL, MA, MD, ME, MI, MN, MO, MT, NC, NE, NH, NY, RI, SD, VA, VT, WI, WY.

GRADE 8 MATH, PERCENTAGES PROFICIENT/BASIC

Although trends in middle school math were somewhat more variable than those in elementary school math, our analyses nevertheless found consistency between state tests and NAEP in the general direction of trends in percentages proficient/basic. Of the 30 states with gains since 2002 on their state middle school (usually grade 8) math test, 28 also showed gains on the NAEP grade 8 math test between 2003 and 2007. However, the NAEP gains were slight in 15 of these states, while the state test gains were moderate-to-large in all but 2 states. As illustrated in **table 4-I**, only 1 state reported a slight decline on NAEP.

Table 4-I. Comparison of Percentage Proficient/Basic Trends on State Tests and NAEP in Middle School/Grade 8 Math

	Type of Trend	NAEP percentage basic: ↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	Total
State test percentage proficient	↑ Moderate-to-large gain	↑ ↑ 12 states: AR, FL, GA, LA, MA, MD, MS, NJ, NM, PA, TN, TX	↑ ↗ 14 states: AK, AL, CO, HI, IA, ID, IN, KY, NV, OH, OK, OR, UT, WA	↑ ✪ 1 state: NE	↑ ↘ 1 state: WV		28
	↗ Slight gain	↗ ↑ 1 state: ND	↗ ↗ 1 state: SC				2
	✪ No change	✪ ↑ 1 state: AZ					1
	↘ Slight decline		↘ ↗ 1 state: CA				1
	↓ Moderate-to-large decline			↗ ✪ 1 state: MT			1
	Total	14	16	2	1	0	33

Table reads: Twenty-eight states demonstrated moderate-to-large gains on state middle school math tests in the post-NCLB years with comparable data. Of these, 12 states also showed moderate-to-large gains on NAEP grade 8 math tests between 2003 and 2007, while 14 states made slight gains on NAEP, 1 state showed no net change on NAEP, and 1 state had a slight decline.

Note: The following 17 states were not included in this comparison with NAEP because data from state tests were unavailable or covered too few years to constitute a trend: CT, DE, IL, KS, ME, MI, MN, MO, NC, NH, NY, RI, SD, VA, VT, WI, WY.

GRADE 8 MATH, EFFECT SIZES

As **table 4-J** indicates, overall trends in effect sizes were also positive for middle school math, with similar patterns between state tests and NAEP. Of the 29 states with sufficient data for this comparison, 28 demonstrated gains on their state middle school math test, and 26 also showed gains on the NAEP grade 8 math test. In most cases, these gains were moderate-to-large on both assessments. One state had a slight decline on the state test, and one state showed a slight decline on NAEP.

Table 4-J. Comparison of Effect Size Trends on State Tests and NAEP in Middle School/Grade 8 Math

State test effect size: Type of Trend	NAEP effect size: ↑ Moderate-to-large gain	↗ Slight gain	✪ No change	↘ Slight decline	↓ Moderate-to-large decline	Total
↑ Moderate-to-large gain	↑ ↑ 22 states: AR, CA, CO, DE, FL, GA, HI, ID, IN, KS, KY, LA, MS, NJ, NM, OH, OK, OR, PA, TN, TX, WA	↑ ↗ 1 state: IA	↑ ✪ 1 state: UT	↑ ↘ 1 state: WV		25
↗ Slight gain	↗ ↑ 3 states: ND, NV, SC					3
✪ No change						0
↘ Slight decline	↘ ↑ 1 state: AZ					1
↓ Moderate-to-large decline						0
Total	26	1	1	1		29

Table reads: Twenty-five states demonstrated moderate-to-large average annual gains in effect size on state middle school math tests in the post-NCLB years with comparable data. Of these, 22 states also showed moderate-to-large average annual gains in effect size on NAEP grade 8 math tests between 2003 and 2007, while 1 state made a slight gain on NAEP, 1 state showed no net change on NAEP, and 1 state had a slight decline.

Note: The following 21 states were not included in this comparison with NAEP because data from state tests were unavailable or covered too few years to constitute a trend: AK, AL, CT, IL, MA, MD, ME, MI, MN, MO, MT, NC, NE, NH, NY, RI, SD, VA, VT, WI, WY.

Magnitude of Overall Achievement Gains on State Tests and NAEP

In addition to comparing the direction of trends on state tests and NAEP, we compared the size of achievement changes on both assessments. Specifically, we looked at whether average yearly changes on state tests were larger than average yearly changes on NAEP, in terms of both percentages proficient/basic and effect sizes. Like the other analyses in this report, this comparison included only states with three or more years of comparable state test data.

Table 4-K displays median percentage point gains in the percentage proficient/basic on state tests and NAEP, as well as the median gains in effect sizes, expressed in standard deviation units. The table also shows the minimum gain (which in most cases means a decline) in the state with the least improvement and the maximum gain in the state with the most improvement. As the table shows, median gains on state tests were larger than median gains on the corresponding NAEP tests across all four grade and subject combinations and on both indicators.

Table 4-K. Size of Achievement Gains on State Tests and NAEP

	Average Annual Gains in Percentages Proficient/Basic							
	Elementary Reading		Middle School Reading		Elementary Math		Middle School Math	
	State test	NAEP	State test	NAEP	State test	NAEP	State test	NAEP
Median	1.8	1.0	1.6	0.2	2.6	1.4	2.0	0.8
Minimum	0.2	-0.6	-1.0	-1.3	0.3	-0.1	-1.3	-0.3
Maximum	5.0	2.6	7.0	1.2	7.7	2.7	7.5	2.3
	Average Annual Gains in Effect Sizes							
	Elementary Reading		Middle School Reading		Elementary Math		Middle School Math	
	State test	NAEP	State test	NAEP	State test	NAEP	State test	NAEP
Median	0.04	0.02	0.03	0.00	0.06	0.05	0.05	0.03
Minimum	-0.02	-0.03	-0.03	-0.04	-0.03	0.00	-0.01	-0.01
Maximum	0.09	0.06	0.22	0.03	0.23	0.10	0.34	0.08

Table reads: On state elementary school reading tests, the median average yearly gain in the percentage proficient at the elementary grade level analyzed was 1.8 percentage points per year for the post-NCLB years with comparable data. The median average annual gain in the percentage basic on the NAEP grade 4 reading test was 1.0 percentage points.

Note: The numbers in the top half of the table are expressed in terms of percentage points and represent average yearly gains in percentages of students scoring at the proficient level on state tests and the basic level on NAEP. The numbers in the bottom half of the table are expressed in standard deviation units and represent average yearly gains in effect sizes. Negative numbers represent a decline in achievement.

Table 4-L provides another view of the magnitude of gains by displaying the number of states with greater gains on their own tests than on NAEP and vice versa. We found substantially more states with larger gains on state tests than on NAEP, in terms of both percentages proficient/basic and effect sizes, but we did find a notable number of states with larger gains on NAEP than on state tests. Of the 252 comparisons for which sufficient data were available, gains were larger on state tests in 188 cases (75%) and larger on NAEP in 55 cases (22%). In nine states, the gains on both assessments were equal.

Table 4-L. Number of States with Larger Gains on State Tests or on NAEP

Comparison	State Larger	NAEP Larger	NAEP & State Equal
	Grade 4 Reading		
Percentage Proficient	25	8	1
Effect Size	19	8	3
	Grade 8 Reading		
Percentage Proficient	27	7	0
Effect Size	24	5	1
	Grade 4 Math		
Percentage Proficient	25	8	0
Effect Size	19	7	3
	Grade 8 Math		
Percentage Proficient	26	6	1
Effect Size	23	6	0
Total	188	55	9

Table reads: When average yearly gains in percentages of students scoring at or above the proficient level on state elementary reading tests were compared with average annual gains in the percentage scoring at or above the basic level on NAEP grade 4 reading tests, the gains were larger on the state test in 25 states and were larger on NAEP in 8 states; gains were the same size on both assessments in 1 state.

Note: The numbers of states in each row do not total 50 because some states had insufficient state test data to make comparisons for that subject or grade level.

Possible Explanations for Similarities and Differences on NAEP and State Tests

Several studies have noted the difference between NAEP and state test results, and some have cited these differences as a reason to question whether student achievement is really improving (see, for example, Cronin et al., 2007; Peterson & Hess, 2006; Fuller et al., 2006; Hall & Kennedy, 2006; Achieve, n.d.; Lee, 2006). In our 2007 version of this achievement study, we found little relationship between NAEP and state test results, except in grade 4 reading. This year, unlike last year, we had data from the 2007 NAEP administration, which meant that we could look at NAEP trends over a four-year period instead of a two-year period.

This year's analysis found a fair amount of consistency in the direction of trends on NAEP and state tests. The fact that learning gains are appearing on measures in addition to state tests suggests that students may be learning more reading and mathematics and that some gains on state tests may have "generalized" to NAEP.

Although NAEP and state tests have often moved in the same upward direction, they have not done so to the same degree in many cases. Gains on state tests still tend to be larger than gains on NAEP. Several possible factors may explain these differences in the magnitude of gains:

- **Instruction is more closely aligned to state standards than NAEP frameworks.** State tests used for No Child Left Behind must be aligned to the state's academic content standards. Ideally, the content of tests should match what is taught in most classrooms. However, NAEP is not aligned to any particular state's standards and therefore may not be instructionally sensitive—in other words, it may not reflect what students are actually being taught in class. State tests may be more instructionally sensitive than NAEP, and therefore reflect larger gains.

- **Score inflation on state tests.** Many researchers believe that scores on state tests have become inflated as a result of inappropriate teaching to state tests (e.g., Koretz, 2005). The more teachers engage in narrow test preparation targeted at the specific format and content of state tests, the more likely it is that state test scores will increase without real, meaningful gains in students' broader knowledge of reading and math. This is a less optimistic explanation for why state test scores may be showing greater gains than NAEP.

- **Motivation.** The state tests used for No Child Left Behind have high stakes for educators and somewhat high stakes for students. Federal and state sanctions or incentives for districts and schools are determined largely by the results of state tests. And in some states, the results of state tests are also used to determine whether students will graduate or be promoted. State test scores are reported to parents and published in the media and can be easily found online. To avoid the sanctions and negative publicity that accompany low test scores, teachers and administrators often go to great lengths to encourage students to take these tests seriously. NAEP, by contrast, has no stakes for administrators, teachers, or students because it is not connected to any rewards, sanctions, or future outcomes for them. Neither students nor their parents receive any individual NAEP results. For these reasons, students may not be motivated to perform their best on NAEP.

- **Some differences in grades and years analyzed.** The grades and time spans we compared for state tests and NAEP were not always identical. As explained earlier, for a small number of states we substituted another grade if state test data for grade 4 or 8 were not available. In addition, while the time frame for NAEP data was 2003 to 2007, the state test score trends covered various years between 2002 and 2007, depending on the data available from each state. These differences may yield different trends.

The reasons for the greater discrepancies between NAEP and state tests in middle school reading are not readily apparent.

It is likely that some combination of the above factors explain the upward trends on both state tests and NAEP and variations in the magnitude of gains. Different factors may be present in various states and at different grade levels.

CHAPTER 5

Trends in Achievement Gaps on State Tests

Summary of Findings

A primary purpose of the No Child Left Behind Act is to close achievement gaps between different subgroups of students, such as African American and white students, Latino and white students, or students from low-income families and those whose families are not low-income. Thus, a major research question for this study is whether achievement gaps have narrowed since 2002. Here are our main findings about achievement gaps:

- **Achievement gaps on state tests have narrowed since 2002 much more often than they have widened.** When we looked at gaps in percentages proficient in reading and math for various subgroups in all states with sufficient data, we found 327 instances where gaps had narrowed over the years analyzed and only 76 instances where gaps had widened; in 20 instances gaps showed no net change. When we examined gaps in effect sizes, we found that gaps had narrowed in 184 cases and widened in only 56 cases; in 30 cases there was no net change in the gap. The percentage proficient and effect size indicators revealed the same overall trends of narrowing or widening, although percentages proficient generally gave a more positive picture of achievement gap trends than effect sizes.

- **For the African American and low-income subgroups, far more states showed gaps narrowing than widening at all grade levels analyzed in both reading and math.** For example, in elementary reading, gaps between African American and white students narrowed in 13 states in terms of both percentages proficient and effect sizes, while the gap widened on both indicators in only 1 state. The low-income subgroup had similar results in elementary school reading; in 10 states low-income gaps narrowed according to both indicators, and in no state did this gap widen. Instances of gaps narrowing were more prevalent than gaps widening for the African American and low-income subgroups at the upper grade levels as well.

- **Gap trends for the Latino subgroup were largely positive, while trends for the Native American subgroup were more mixed.** Because of data limitations, however, these findings are less conclusive than the findings for other subgroups. In many states, the Latino and Native American subgroups were not included in our analyses because they were too small or had changed significantly in size in recent years.

- **Gap trends in the states with the largest subgroup enrollments generally reflected the national findings.** When we examined gap trends in the states that together accounted for more than half of the total enrollment of a specific subgroup nationwide, we found the trends were consistent with the national pattern of gaps on state tests narrowing much more often than they widened. For example, five states (California, Texas, Florida, New York, and Arizona) account for approximately 70% of the Latino student enrollment nationally. Four of these states had sufficient data, and in these states, Latino-white achievement gaps narrowed more often than they widened. Therefore, the gap findings do not appear to be an anomaly attributable to unusual progress in states with small Latino populations. The same was true for the other three subgroups studied.

How We Analyzed Trends in Achievement Gaps on State Tests

Our study focused on gaps between the white subgroup and three racial/ethnic subgroups—African American, Latino, and Native American students—as well as gaps between low-income and non-low-income students. Although we collected and reported detailed test data for Asian students in all of the online state profiles developed for this report (www.cep-dc.org), we did not include this subgroup in our national study because Asian students generally perform as well as or better than white students in all but a few states. And although we collected and reported detailed test data for students with disabilities and English language learners in all of the state profiles, we did not analyze trends for these two subgroups in our national study. As explained in chapter 2, this is because the comparability of test results for students with disabilities and ELLs has been affected by a variety of policy changes since 2002 governing which of these students are tested for NCLB accountability, how they are tested, and when their scores are counted as proficient.

We analyzed gaps in reading and math from various perspectives, using two indicators of achievement—the percentage of students scoring at or above the proficient level and effect size. To determine whether states made progress on achievement gaps, we calculated average yearly changes in gaps. The detailed data used to compute these average yearly changes can be viewed in the online state profiles.

Three issues related to subgroups raised particular concerns that had to be addressed. First, some subgroups were small relative to the total number of test-takers in certain states. Second, some subgroups had changed significantly in size in some states over the years analyzed. Third, the percentages of students in each subgroup varied widely across states.

The first two issues—small subgroups and changing subgroups—are concerns because they can produce volatility in test scores that can complicate judgments about gap trends and lead to conclusions that may not be reliable. For the main nationwide analyses, for which we wanted to make a firm judgment about trends, we excluded small subgroups (those with fewer than 500 test-takers) and changing subgroups (those that had grown or shrunk by 25% over the period analyzed), as explained in chapter 2. Applying these rules meant that for some subgroups, we had to exclude a large number of states from the analyses. This raised another issue—namely, whether the picture of gap trends that emerged from the remaining states was indicative of what was happening nationally. Therefore, for some of our analyses, we separately report the findings for the small and changing subgroups, while noting that these results should be viewed with caution.

We also considered the third issue, the distribution of subgroups across states. As we did with the overall achievement trends in chapter 3, we present the gap trends here in terms of numbers of states showing various trends without noting how many students in a given subgroup are enrolled in each state. To address the distribution concern, for each subgroup we performed an additional analysis of trends in states with the highest enrollments of that subgroup. We checked to see whether the broad trends in achievement gaps in the high-enrollment states were consistent with nationwide patterns and found that they were.

In this chapter, we first present a broad picture of gap trends nationwide. Second, we examine specific trends for each subgroup, by subject and by grade level. Third, we look at the size of the gaps. Then we offer possible explanations for the gap trends we identified. **Tables 5-O** and **5-P** at the end of this chapter provide a state-by-state summary of achievement gap trends in reading and math.

Instances of Gaps Narrowing and Widening

To get an overall picture of gap trends, we looked at trends for four subgroups in each state (African American, Latino, Native American, and low-income students), at three grade levels (elementary, middle, and high school), in two subjects (reading and math), and on two indicators (percentages proficient and effect sizes). Each of these data points we refer to as an "instance." For example, the gap in the percentage proficient between African American and white students in middle school reading in Florida is counted as one instance. The same gap in Florida at the high school level is counted as another instance. We tallied the numbers of specific gaps that have narrowed, widened, and shown no net change over the years analyzed. These counts are derived from the state-by-state summary tables 5-O and 5-P at the end of this chapter. Instances where data were unavailable, or where subgroups were small or changing, were not counted in this analysis.

The results of this analysis are displayed in **table 5-A**. The table indicates that since 2002, the number of instances of achievement gaps that have narrowed far exceeds the number that have widened or stayed the same. Of all of the instances analyzed using the percentage proficient indicator, 77% showed gaps narrowing while 18% showed gaps widening. Using the effect size indicator, 68% of the instances showed gaps narrowing and 21% showed gaps widening.

Table 5-A. Summary of Gap Trends on State Tests Since 2002

	Number (*percentage*) of			
	→← Gaps that narrowed	✪ Gaps with no change	←→ Gaps that widened	Total # of instances
Percentage proficient	327 *(77%)*	20 *(5%)*	76 *(18%)*	423
Effect size	184 *(68%)*	30 *(11%)*	56 *(21%)*	270

Table reads: Of all the gaps analyzed in all states for four subgroups, at three grade levels, in two subjects, and by two indicators, 327 gaps narrowed in terms of percentages proficient. This is equal to 77% of the total number of gaps analyzed.

As explained in chapter 2, the percentage proficient indicator, the primary indicator used by NCLB, has some limitations. With gap analyses in particular, the location of the proficient cut score on a state test can make gaps appear narrower because of the statistical anomaly explained in appendix 2. In part to address this problem, we used effect sizes, which are based on mean test scores, as an additional indicator of student achievement. We compared trends on the two indicators for states that provided both types of data (again, excluding instances where the subgroups were small or changing). The results are displayed in **table 5-B**. Both indicators give a positive picture of gap trends, but effect sizes show fewer instances of gaps narrowing than the percentage proficient indicator and more instances of gaps widening or staying the same. In other words, the percentage proficient indicator gives a somewhat rosier picture of trends in achievement gaps.

Table 5-B. Summary of Gap Trends on State Tests for States and Grade Levels with Both Percentage Proficient and Effect Size Data

	Number (*percentage*) of			
	→← Gaps that narrowed	✪ Gaps with no change	←→ Gaps that widened	Total # of instances
Percentage proficient	198 *(78%)*	14 *(5%)*	43 *(17%)*	255
Effect size	174 *(68%)*	25 *(10%)*	56 *(22%)*	255

Table reads: Of all the gaps analyzed in states with sufficient data on both percentages proficient and effect sizes, 198 gaps narrowed in terms of percentage proficient. This is equal to 78% of the total number of gaps analyzed.

States with Across-the-Board Improvements

Another way to view the big picture is to see how many states showed consistent gap trends across all grade levels analyzed. For this analysis, we examined states according to stringent criteria—they had to show trends in achievement gaps consistently at all three grade levels in terms of both percentages proficient and effect sizes. They also had to have at least three years of comparable test data and subgroups that were not too small or had not changed significantly in size. In short, these were states with adequate data and stable subgroup sizes in which gaps narrowed at all three grades analyzed—one grade at the elementary, middle, and high school levels. Most states did not meet these criteria because their gaps narrowed at one or two grade levels, but not all three. But some states did demonstrate across-the-board improvements:

- In five states (Iowa, Louisiana, Mississippi, Texas, and West Virginia), achievement gaps between African American and white students have narrowed in reading since 2002 at all three grade levels analyzed on both indicators. In six states (Indiana, Iowa, Louisiana, Mississippi, Tennessee, and West Virginia) the African American-white gap narrowed in math according to the same criteria.

- In Texas, achievement gaps in math narrowed between Latino and white students and between low-income and non-low-income students at all three grade levels analyzed on both indicators.

- No state showed any gaps widening across all grade levels on both indicators.

Achievement Gaps by Subgroup, Subject, and Grade Level

To probe gap trends in more detail, we did separate analyses in reading and math for each subgroup at each of the three grade levels. These analyses included states with both percentages proficient and effect sizes, as well as states with data for just one indicator. As explained in chapter 2, we examined one grade each at the elementary, middle, and high school levels in every state—usually grades 4, 8, and 10 but sometimes other grades when comparable data were not available.

AFRICAN AMERICAN STUDENTS, READING

Table 5-C depicts several trends in gaps between African American and white students in reading. At the **elementary** level, far more states showed gaps narrowing between African American and white students than widening. In 13 states, gaps narrowed on both the percentage proficient and effect size indicators. In 1 state, the gap widened on both indicators. Of the states with percentage proficient data only, gaps narrowed in 5 states and widened in 1. The results were similar for **middle school**. At the **high school** level, the African American-white gap in reading shrunk on both indicators in 6 states and widened in 2. Of the states with percentage proficient data only, 5 had narrowing gaps and 2 had widening gaps. Fewer states had trends at the high school level than at other grade levels because many states either lacked sufficient high school data or had African American subgroups that were too small or had changed significantly in size.

Across the three grade levels, there were eight instances where states showed contradictory results—gaps narrowed on one indicator and widened on the other. For example, in North Carolina, the elementary reading achievement gap between African American and white students narrowed in terms of the percentage proficient but widened in terms of effect size. The state's online profile reveals that the African American-white gap in the percentage proficient decreased from 16.8 percentage points to 14.9 percentage points in 2007. However, the effect size gap increased from 0.77 standard deviations in 2003 to 0.8 standard deviations in 2007. This small increase, which is not unusual, illustrates how the two indicators might show different trends. One possible explanation is that, in either the African American or white subgroup or both, the performance of students above or below the proficient level might have changed without adding to the percentage proficient. For example, if gains were made by high-performing white students—those who were already scoring above the proficient level—then mean test scores for the white subgroup would go up without a corresponding increase in the percentage proficient. Another possible explanation is that the size of the percentage proficient gap is being affected by the location of the proficient cut score on the state's grade 4 reading test, as explained in appendix 2. These explanations are speculative; a more detailed examination of North Carolina's test scores would be necessary to get the full answer. In any case, having both indicators in agreement provides stronger evidence of achievement gaps actually closing than one indicator alone.

Table 5-C. Gap Trends for African American Students in Reading Since 2002

Type of Trend		Grade	Effect size: Narrowed	No change	Widened	Subgroup small or changed	Data unavailable or too few years
Percentage of students scoring at or above the proficient level	Narrowed	E	13 states: AZ, CO, IN, IA, KY, LA, MS, NV, NM, PA, TX, WA, WV	1 state: FL	3 states: NC, OR[1], TN		5 states: AL, MD, MA, NJ, OK
		M	12 states: FL, IN, IA, KY, LA, MS, NV, SC, TN, TX, WA, WV	1 state: NC	3 states: AR, NM, OR[1]		5 states: AL, MD, NJ, OH, OK
		H	6 states: IA, LA, MS, TN, TX, WV	1 state: SC			5 states: CA, MD, MA, NJ, OK
	No change	E	1 state: SC	1 state: CA	1 state: AR		
		M					
		H			1 state: IN		
	Widened	E			1 state: UT		1 state: OH
		M	1 state: AZ	1 state: CA	1 state: CO		1 state: MA
		H	1 state: KY		2 states: AR, CO		2 states: AL, RI

Continued on page 71

Table 5-C. Gap Trends for African American Students in Reading Since 2002 (continued)

Type of Trend	Grade	Effect size: Narrowed (→←)	No change (✪)	Widened (←→)	Subgroup small or changed (✧)	Data unavailable or too few years (♦/❍)
Percentage of students scoring at or above the proficient level — Subgroup small or changed (✧)	E				✧ ✧ 3 states: HI, ID, ND	✧ ♦/❍ 3 states: AK, MT, SD
	M				✧ ✧ 5 states: HI, ID, ND, PA, UT	✧ ♦/❍ 3 states: AK, MT, SD
	H				✧ ✧ 11 states: AZ, CT, FL, HI, ID, NV, ND, OR, PA, UT, WA	✧ ♦/❍ 6 states: ME, MT, NH, NM, OH, SD
Percentage of students scoring at or above the proficient level — Data unavailable or too few years (♦/❍)	E	♦/❍ →← 1 state: DE				♦/❍ ♦/❍ 16 states: CT, GA, IL, KS, ME, MI, MN, MO, NE, NH, NY, RI, VT, VA, WI, WY
	M				♦/❍ ✧ 1 state: DE	♦/❍ ♦/❍ 16 states: CT, GA, IL, KS, ME, MI, MN, MO, NE, NH, NY, RI, VT, VA, WI, WY
	H	♦/❍ →← 1 state: DE				♦/❍ ♦/❍ 14 states: AK, GA, IL, KS, MI, MN, MO, NE, NY, NC, VT, VA, WI, WY

Table reads: Since NCLB took effect in 2002, reading achievement gaps between African American and white students narrowed at the elementary grade analyzed in 13 states, according to both the percentages of students scoring proficient and effect sizes. In 3 states, the African American-white reading gap narrowed in terms of percentages proficient but widened in terms of effect sizes.

[1] The percentage proficient data covered a different range of years than the effect size data, which may explain the contradictory trends.

AFRICAN AMERICAN STUDENTS, MATH

Trends in the gap between African American and white students in math, displayed in **table 5-D**, are similar to those for reading. At the **elementary level**, the African American-white gap in math narrowed on both the percentage proficient and effect size indicators in 14 states and widened on both in 1 state. In the 8 states with data for one indicator only, the gap shrunk. In **middle school**, the African American-white gap in math narrowed on both indicators in 9 states and widened on both in 1 state. Of the states with percentage proficient data only, 4 states showed gaps narrowing and 2 states showed gaps widening. At the **high school** level, the African American-white math gap shrunk on both indicators in 8 states and widened in none. For the states with percentage proficient data only, gaps narrowed in 6 states and widened in 1. Across grades, there were 7 instances of gaps narrowing on one indicator and widening on the other.

Table 5-D. Gap Trends for African American Students in Math Since 2002

Type of Trend		Grade	Effect size: →← Narrowed	Effect size: ✪ No change	Effect size: ↔ Widened	Effect size: ✧ Subgroup small or changed	Effect size: ♦/❍ Data unavailable or too few years
Percentage of students scoring at or above the proficient level	Narrowed	E	→← →← 14 states: AZ, CA, CO, FL, IN, IA, KY, LA, MS, NM, PA, TN, TX, WV		→← ↔ 1 state: OR[1]		→← ♦/❍ 7 states: AL, GA, MD, MA, NJ, OH, OK
		M	→← →← 9 states: AZ, FL, IN, IA, LA, MS, TN, TX, WV	→← ✪ 2 states: NV, OR	→← ↔ 2 states: CO, UT		→← ♦/❍ 4 states: AL, NJ, OH, OK
		H	→← →← 8 states: CO, IN, IA, LA, MS, SC, TN, WV	→← ✪ 1 state: TX			→← ♦/❍ 6 states: AL, CA, MA, NJ, OH, RI
	No change	E	✪ →← 1 state: NV				
		M					
		H					
	Widened	E	↔ →← 1 state: WA	↔ ✪ 2 states: AR, SC	↔ ↔ 1 state: UT		
		M	↔ →← 3 states: NM, SC, WA	↔ ✪ 1 state: KY	↔ ↔ 1 state: AR		↔ ♦/❍ 2 states: MD, MA
		H					↔ ♦/❍ 1 state: OK

Continued on page 73

Table 5-D. Gap Trends for African American Students in Math Since 2002 (continued)

Type of Trend	Grade	Effect size: ➔➔ Narrowed	✪ No change	↔ Widened	✧ Subgroup small or changed	♦/❍ Data unavailable or too few years
Percentage of students scoring at or above the proficient level: ✧ Subgroup small or changed	E				✧ ✧ 3 states: HI, ID, ND	✧ ♦/❍ 2 states: AK, MT
	M				✧ ✧ 4 states: HI, ID, ND, PA	✧ ♦/❍ 4 states: AK, CA, GA, MT
	H				✧ ✧ 13 states: AZ, AR, CT, FL, HI, ID, KY, NV, ND, OR, PA, UT, WA	✧ ♦/❍ 5 states: GA, ME, MT, NH, NM
Percentage of students scoring at or above the proficient level: ♦/❍ Data unavailable or too few years	E	♦/❍ ➔➔ 1 state: DE				♦/❍ ♦/❍ 17 states: CT, IL, KS, ME, MI, MN, MO, NE, NH, NY, NC, RI, SD, VT, VA, WI, WY
	M				♦/❍ ✧ 1 state: DE	♦/❍ ♦/❍ 17 states: CT, IL, KS, ME, MI, MN, MO, NE, NH, NY, NC, RI, SD, VT, VA, WI, WY
	H	♦/❍ ➔➔ 1 state: DE				♦/❍ ♦/❍ 15 states: AK, IL, KS, MD, MI, MN, MO, NE, NY, NC, SD, VT, VA, WI, WY

Table reads: Since NCLB took effect in 2002, math achievement gaps between African American and white students narrowed at the elementary grade analyzed in 14 states, according to both the percentages of students scoring proficient and effect sizes. In one state, the African American-white math gap narrowed in terms of percentages proficient but widened in terms of effect sizes.

[1] The percentage proficient data covered a different range of years than the effect size data, which may explain the contradictory trends.

SMALL AND CHANGING AFRICAN AMERICAN SUBGROUPS

As the above tables reveal, several states were excluded from our main analyses due to small or changing African American subgroups. For example, as table 5-D above shows, the African American subgroup was small or changing in 5 states at the elementary level and in 18 states at the high school level. We examined gap trends in the excluded states to see whether the nationwide picture would change if they had been included. We found that in states with small and changing African American subgroups, reading and math gap trends showed the same pattern of more gaps narrowing than widening, but the pattern was less pronounced. In reading, gaps narrowed on both indicators across the three grade levels analyzed in 8 instances and widened on both in 5 instances. In math, 7 gaps narrowed on both while 3 gaps widened.

STATES WITH THE LARGEST AFRICAN AMERICAN SUBGROUPS

According to the Common Core of Data from the National Center for Education Statistics (2005), there were 8,376,855 African American (non-Hispanic) students enrolled in public schools in 2005, the latest year for which data are available. Eight states accounted for just over half (4,199,928) of total African American enrollment; these were, in order, Texas, Florida, Georgia, New York, California, North Carolina, Illinois, and Michigan. We examined gap trends in four of these states (Texas, Florida, California, and North Carolina); the rest did not have the required data for this analysis. The results reflected the nationwide picture—instances of gaps narrowing were much more frequent than instances of gaps widening:

- The two states with the largest African American enrollments, Texas and Florida, had very positive results. Gaps narrowed at almost all grade levels in both subjects on both indicators, and there were no instances of gaps widening. At the high school level in Florida, achievement gaps narrowed but this finding should be interpreted with caution because the subgroup changed significantly in size.

- In California and North Carolina, instances of gaps narrowing also outweighed instances of gaps widening, but the overall picture was more mixed. There were more cases of gaps staying the same, and in each state one grade level had a gap that widened on one indicator (middle school in California, elementary school in North Carolina).

LATINO STUDENTS, READING

While more states showed gaps in reading between Latino and white students narrowing than widening, the analysis is limited because many states lacked sufficient comparable data or had Latino subgroups that were too small or had changed significantly in size (see **table 5-E**). At the **elementary** level, gaps between Latino and white students narrowed in 7 states on both percentages proficient and effect sizes. In no state did gaps widen on both. Gaps also narrowed in all 6 states with percentage proficient data only. In **middle school**, the Latino-white reading gap narrowed on both indicators in 3 states and widened in 1 state. In the 4 states with percentage proficient data only, gaps narrowed. At the **high school** level, just 3 states had sufficient data for analyses of both indicators. In 1 of these states, gaps narrowed on both indicators, while in the other 2 states, gap trends were mixed. Of the states with percentage proficient data only, gaps narrowed in 3 states and widened in 2.

Table 5-E. Gap Trends for Latino Students in Reading Since 2002

Type of Trend		Grade	Effect size: Narrowed	Effect size: No change	Effect size: Widened	Effect size: Subgroup small or changed	Effect size: Data unavailable or too few years
Percentage of students scoring at or above the proficient level	Narrowed	E	7 states: AZ, CA, FL, ID, IA, NV, NM	1 state: CO	2 states: TX, UT[1]		6 states: AL, MD, MA, NJ, OH, OK
		M	3 states: IA, NM, TX	1 state: CO	2 states: ID[1], UT[1]		4 states: AL, MA, OH, OK
		H	1 state: TX		2 states: LA, UT[1]		3 states: AL, MA, NJ
	No change	E					
		M	1 state: AZ				
		H					2 states: CA, OK
	Widened	E					
		M			1 state: CA		
		H					2 states: NM, RI

Continued on page 76

Table 5-E. Gap Trends for Latino Students in Reading Since 2002 (continued)

Type of Trend	Grade	Effect size: ➔➔ Narrowed	✪ No change	⟷ Widened	✧ Subgroup small or changed	♦/❍ Data unavailable or too few years
Percentage of students scoring at or above the proficient level: ✧ Subgroup small or changed	E				✧ ✧ **13 states:** AR, HI, IN, KY, LA, MS, NC, ND, OR, PA, SC, TN, WA	✧ ♦/❍ **4 states:** AK, MT, SD, WV
	M				✧ ✧ **15 states:** AR, FL, HI, IN, KY, LA, MS, NV, NC, ND, OR, PA, SC, TN, WA	✧ ♦/❍ **6 states:** AK, MD, MT, NJ, SD, WV
	H				✧ ✧ **18 states:** AZ, AR, CO, CT, FL, HI, ID, IN, IA, KY, MS, NV, ND, OR, PA, SC, TN, WA	✧ ♦/❍ **7 states:** ME, MD,MT, NH, OH, SD, WV
Percentage of students scoring at or above the proficient level: ♦/❍ Data unavailable or too few years	E				♦/❍ ✧ **1 state:** DE	♦/❍ ♦/❍ **16 states:** CT, GA, IL, KS, ME, MI, MN, MO, NE, NH, NY, RI, VT, VA, WI, WY
	M				♦/❍ ✧ **1 state:** DE	♦/❍ ♦/❍ **16 states:** CT, GA, IL, KS, ME, MI, MN, MO, NE, NH, NY, RI, VT, VA, WI, WY
	H				♦/❍ ✧ **1 state:** DE	♦/❍ ♦/❍ **14 states:** AK, GA, IL, KS, MI, MN, MO, NE, NY, NC, VT, VA, WI, WY

Table reads: Since NCLB took effect in 2002, reading achievement gaps between Latino and white students narrowed at the elementary grade analyzed in seven states, according to both the percentages of students scoring proficient and effect sizes. In two states, the Latino-white reading gap narrowed in terms of percentages proficient but widened in terms of effect sizes.

[1] The percentage proficient data covered a different range of years than the effect size data, which may explain the contradictory trends.

LATINO STUDENTS, MATH

In the limited number of states with sufficient data and stable subgroups, the Latino-white gap in math narrowed at the **elementary** level in 8 states, according to both percentages proficient and effect sizes. In no state did the gaps widen on both indicators (see **table 5-F**). In all 6 states with percentage proficient data only, gaps shrunk. At the **middle school** level, the Latino-white math gap narrowed on both indicators in 5 states and widened on both indicators in none. In the 4 states with percentage proficient data only, this gap narrowed in 2 and widened in 2. At the **high school** level, only 3 states had sufficient data on both indicators and subgroups that met our criteria. The Latino-white math gap narrowed on both indicators in 1 state, widened in another, and showed mixed results in the third. Of the states with percentage proficient data only, gaps narrowed in 6 states and widened in 1 state.

Table 5-F. Gap Trends for Latino Students in Math Since 2002

Type of Trend		Grade	Effect size: Narrowed	Effect size: No change	Effect size: Widened	Effect size: Subgroup small or changed	Effect size: Data unavailable or too few years
Percentage of students scoring at or above the proficient level	Narrowed	E	8 states: AZ, CA, FL, IA, NV, NM, TX, UT	1 state: CO	1 state: ID[1]		6 states: AL, MD, MA, NJ, OH, OK
		M	5 states: AZ, CO, ID, IA, TX		1 state: UT[1]		2 states: OH, OK
		H	1 state: TX	1 state: LA			6 states: AL, CA, MA, NJ, NM, RI
	No change	E					
		M					
		H					
	Widened	E					
		M	1 state: NM				2 state: AL, MA
		H			1 state: UT		1 state: OK

Continued on page 78

Table 5-F. Gap Trends for Latino Students in Math Since 2002 (continued)

Type of Trend	Grade	Effect size: →← Narrowed	✪ No change	←→ Widened	✧ Subgroup small or changed	♦/❍ Data unavailable or too few years
Percentage of students scoring at or above the proficient level: ✧ Subgroup small or changed	E				✧ ✧ 12 states: AR, HI, IN, KY, LA, MS, ND, OR, PA, SC, TN, WA	✧ ♦/❍ 4 states: AK, GA, MT, WV
	M				✧ ✧ 14 states: AR, FL, HI, IN, KY, LA, MS, NV, ND, OR, PA, SC, TN, WA	✧ ♦/❍ 7 states: AK, CA, GA, MD, MT, NJ, WV
	H				✧ ✧ 18 states: AZ, AR, CO, CT, FL, HI, ID, IN, IA, KY, MS, NV, ND, OR, PA, SC, TN, WA	✧ ♦/❍ 6 states: GA, ME, MT, NH, OH, WV
Percentage of students scoring at or above the proficient level: ♦/❍ Data unavailable or too few years	E				♦/❍ ✧ 1 state: DE	♦/❍ ♦/❍ 17 states: CT, IL, KS, ME, MI, MN, MO, NE, NH, NY, NC, RI, SD, VT, VA, WI, WY
	M				♦/❍ ✧ 1 state: DE	♦/❍ ♦/❍ 17 states: CT, IL, KS, ME, MI, MN, MO, NE, NH, NY, NC, RI, SD, VT, VA, WI, WY
	H				♦/❍ ✧ 1 state: DE	♦/❍ ♦/❍ 15 states: AK, IL, KS, MD, MI, MN, MO, NE, NY, NC, SD, VT, VA, WI, WY

Table reads: Since NCLB took effect in 2002, math achievement gaps between Latino and white students narrowed at the elementary grade analyzed in eight states, according to both the percentages of students scoring proficient and effect sizes. In one state, the Latino-white math gap narrowed in terms of percentages proficient but widened in terms of effect sizes.

[1] The percentage proficient data covered a different range of years than the effect size data, which may explain the contradictory trends.

SMALL AND CHANGING LATINO SUBGROUPS

Among the four subgroups we analyzed, the Latino subgroup was small or had changed significantly in size in the most states. We examined Latino-white gap trends in the states excluded for this reason. The results are shown in **tables 5-G** and **5-H** but should be viewed with caution because of methodological issues. In reading, the results were less positive than in the states with more stable Latino subgroup populations. For states with both percentage proficient and effect size data, gaps narrowed and widened in an equal number of states at the elementary and middle school levels. At the high school level, results were far more positive; gaps narrowed in 10 states and widened in just 1 state. In math, the results for states with small or changing Latino subgroups (table 5-H) were somewhat less positive than those for the states with stable populations (table 5-F), but they did not conflict with the general findings of this chapter. At the elementary level, more states showed math achievement gaps narrowing than widening, but at the middle school level, results were more mixed.

It is difficult to make judgments about achievement gaps when the composition of the tested subgroup has changed rapidly, as was the case with the Latino subgroup in many states. We analyzed the numbers of Latino test-takers in all states and found that, depending on the tested subject and grade level, the number of states with increasing numbers of Latino test-takers ranged from 33 to 41, while the number of states with decreasing numbers ranged from 1 to 6. NCLB requires states to test nearly all students, including students who are not fluent in English if they have been in a U.S. school for at least a year. The expansion of this pool of tested students can have an effect on achievement gaps that has little to do with actual progress, or lack of progress, in learning.

STATES WITH THE HIGHEST LATINO STUDENT ENROLLMENTS

According to the Common Core of Data (National Center for Education Statistics, 2005), there were 9,641,407 Latino students enrolled in 2005, of which 70% were enrolled in just five states: California, Texas, Florida, New York, and Arizona. The top two states, California and Texas, accounted for more than half (5,052,510) of Latino enrollment nationally. We analyzed four of these states, as New York did not have the required data. In keeping with the findings for all states, instances of achievement gaps narrowing in states with the highest Latino student populations far outnumbered instances of gaps widening.

- In California, gaps between Latino and white students narrowed at the elementary level in reading and math on both indicators. Gaps widened on both indicators in middle school reading; data were limited in math. At the high school level, effect size data were missing, but gaps in percentages proficient showed no change in reading and narrowed in math.

- In Texas, gaps narrowed at all levels in both subjects and on both indicators, except for elementary reading, where the gap widened according to the effect size indicator only.

- Florida and Arizona both showed gaps narrowing at the elementary level. In Arizona, gaps narrowed in middle school as well. Gaps also narrowed at the middle school level in Florida, and at the high school level in both Florida and Arizona, but in these instances the subgroup changed notably in size, so these results should be interpreted with caution.

Table 5-G. Small or Changing Subgroups: Latino-White Gap Trends in Reading

Underlined = Small **Bold** = Changing ***Bold & Italics*** = Both small & changing

Type of Trend	Grade	Effect size: Narrowed (→←)	Effect size: No change (✪)	Effect size: Widened (↔)	Effect size: Data unavailable or too few years (♦/❍)
Percentage of students scoring at or above the proficient level — Narrowed (→←)	E	→← →← 3 states: ***HI***, PA, WA	→← ✪ 1 state: NC	→← ↔ 1 state: MS, ***ND***, OR, TN	→← ♦/❍ 2 states: MT, ***WV***
	M	→← →← 7 states: FL, ***HI***, NV, NC, PA, TN, WA		→← ↔ 1 state: ND	→← ♦/❍ 5 states: AK, **MD**, MT, **NJ**, SD
	H	→← →← 10 states: AR, AZ, FL, ***HI***, ID, IN, IA, NV, ND, WA	→← ✪ 1 state: ***MS***	→← ↔ 5 states: CO, CT, OR, SC, TN	→← ♦/❍ 4 states: ME, MT, **OH**, SD
Percentage of students scoring at or above the proficient level — No change (✪)	E		✪ ✪ 1 state: KY		
	M				
	H		✪ ✪ 1 state: ***PA***		
Percentage of students scoring at or above the proficient level — Widened (↔)	E		↔ ✪ 1 state: SC	↔ ↔ 3 states: AR, IN, LA	↔ ♦/❍ 2 states: AK, SD
	M			↔ ↔ 7 state: AR, IN, KY, LA, ***MS***, OR, SC	↔ ♦/❍ 1 state: ***WV***
	H			↔ ↔ 1 state: KY	↔ ♦/❍ 3 states: **MD**, ***NH***, WV

Continued on page 81

Table 5-G. Small or Changing Subgroups: Latino-White Gap Trends in Reading (continued)

Type of Trend	Grade	Effect size: ➔➔ Narrowed	Effect size: ✪ No change	Effect size: ↔ Widened	Effect size: ♦/❍ Data unavailable or too few years
Percentage at or above the proficient level	E	♦/❍ ➔➔ 1 state: DE			
♦/❍ Data unavailable or too few years	M		♦/❍ ✪ 1 state: DE		
	H	♦/❍ ➔➔ 1 state: DE			

Table reads: Since NCLB took effect in 2002, reading achievement gaps between Latino and white students in states with small or changing subgroups narrowed at the elementary grade analyzed in three states, according to both the percentages of students scoring proficient and effect sizes. In one state, the Latino-white math gap narrowed in terms of percentages proficient but showed no change in terms of effect sizes.

Table 5-H. Small or Changing Subgroups: Latino-White Gap Trends in Math

Underlined = Small **Bold** = Changing ***Bold & Italics*** = Both small & changing

Type of Trend	Grade	Effect size: Narrowed (→←)	Effect size: No change (✪)	Effect size: Widened (↔)	Effect size: Data unavailable or too few years (♦/❍)
Percentage of students scoring at or above the proficient level — Narrowed (→ ←)	E	→← →← 5 states: HI, ***ND,*** OR, PA, TN		→← ↔ 1 state: KY	→← ♦/❍ 2 states: AK, GA
	M	→← →← 5 states: FL, ***HI***, NV, PA, TN	→← ✪ 1 state: OR	→← ↔ 1 state: AR	→← ♦/❍ 5 states: AK, CA, GA, MT, NJ
	H	→← →← 13 states: AZ, CT, FL, ***HI***, ID, IN, IA, NV, ND, OR, PA, SC, TN	→← ✪ 1 state: CO		→← ♦/❍ 3 states: GA, MT, OH
Percentage of students scoring at or above the proficient level — No change (✪)	E		✪ ✪ 1 state: IN		✪ ♦/❍ 1 state: MT
	M				
	H			✪ ↔ 1 state: ***MS***	
Percentage of students scoring at or above the proficient level — Widened (↔)	E	↔ →← 2 states: MS, WA		↔ ↔ 3 states: AR, LA, SC	↔ ♦/❍ 1 state: ***WV***
	M	↔ →← 2 states: ND, WA		↔ ↔ 5 states: IN, KY, LA, ***MS***, SC	↔ ♦/❍ 2 states: MD, ***WV***
	H	↔ →← 1 state: WA	↔ ✪ 1 state: AR	↔ ↔ 1 state: KY	↔ ♦/❍ 3 states: ME, ***NH, WV***

Continued on page 83

Table 5-H. Small or Changing Subgroups: Latino-White Gap Trends in Math (continued)

Type of Trend	Grade	Effect size: ➔➔ Narrowed	✪ No change	↔ Widened	♦/❍ Data unavailable or too few years
Percentage at or above the proficient level	E	♦/❍ ➔➔ 1 state: DE			
♦/❍ Data unavailable or too few years	M	♦/❍ ➔➔ 1 state: DE			
	H	♦/❍ ➔➔ 1 state: DE			

Table reads: Since NCLB took effect in 2002, math achievement gaps between Latino and white students in states with small or changing subgroups narrowed at the elementary grade analyzed in five states, according to both the percentages of students scoring proficient and effect sizes. In one state, the Latino-white math gap narrowed in terms of percentages proficient but widened in terms of effect sizes.

NATIVE AMERICAN STUDENTS, READING

Trends in gaps between Native American and white students in reading were mixed, as displayed in **table 5-I**. Many states lacked sufficient comparable data or had Native American subgroups that were too small to analyze. Few states provided both effect size and percentage proficient data for this subgroup. At the **elementary** level, more states showed gaps narrowing than widening according to percentages proficient. However, several instances could be found of percentage proficient gaps narrowing while effect size gaps widened. At the **middle** and **high school** levels, the results were mixed and no clear pattern emerged.

Table 5-I. Gap Trends for Native American Students in Reading Since 2002

Type of Trend	Grade	Effect size: →← Narrowed	Effect size: ✪ No change	Effect size: ←→ Widened	Effect size: ✧ Subgroup small or changed	Effect size: ♦/❍ Data unavailable or too few years
Percentage of students scoring at or above the proficient level: ↓↑ Narrowed	E	→← →← 3 states: AZ, NM, OR	→← ✪ 1 state: ND	→← ←→ 5 states: CA, NV, NC, UT, WA		→← ♦/❍ 3 states: MT, OK, SD
	M	→← →← 3 states: ND, TX, WA		→← ←→ 1 state: NC		→← ♦/❍ 3 states: AL, MT, OK
	H			→← ←→ 4 states: CO, TX, UT, WA		→← ♦/❍ 4 states: CA, MT, OK, RI
✪ No change	E					
	M			✪ ←→ 1 state: CO		
	H	✪ →← 1 state: AZ				
↕ Widened	E	←→ →← 1 state: CO		←→ ←→ 1 state: TX		←→ ♦/❍ 2 states: AK, FL
	M	←→ →← 2 states: AZ, NM	←→ ✪ 1 state: CA	←→ ←→ 1 state: NV		←→ ♦/❍ 3 states: AK, FL, SD
	H					←→ ♦/❍ 3 states: AL, NM, SD

Continued on page 85

Table 5-I. Gap Trends for Native American Students in Reading Since 2002 (continued)

Type of Trend	Grade	Effect size: →← Narrowed	✪ No change	←→ Widened	✧ Subgroup small or changed	♦/❍ Data unavailable or too few years
Percentage of students scoring at or above the proficient level						
✧ Subgroup small or changed	E				✧ ✧ 9 states: HI, ID, IN, IA, LA, MS, PA, SC, TN	✧ ♦/❍ 7 states: AR, AL, MD, MA, NJ, OH, WV
	M				✧ ✧ 11 states: HI, ID, IN, IA, LA, MS, OR, PA, SC, TN, UT	✧ ♦/❍ 6 states: AR, MD, MA, NJ, OH, WV
	H				✧ ✧ 13 states: CT, HI, ID, IN, IA, LA, MS, NV, ND, OR, PA, SC, TN	✧ ♦/❍ 9 states: AR, FL, ME, MD, MA, NH, NJ, OH, WV
♦/❍ Data unavailable or too few years	E				♦/❍ ✧ 1 state: DE	♦/❍ ♦/❍ 17 states: CT, GA, IL, KS, KY, ME, MI, MN, MO, NE, NH, NY, RI, VT, VA, WI, WY
	M				♦/❍ ✧ 1 state: DE	♦/❍ ♦/❍ 17 states: CT, GA, IL, KS, KY, ME, MI, MN, MO, NE, NH, NY, RI, VT, VA, WI, WY
	H				♦/❍ ✧ 1 state: DE	♦/❍ ♦/❍ 15 states: AK, GA, IL, KS, KY, MI, MN, MO, NE, NY, NC, VT, VA, WI, WY

Table reads: Since NCLB took effect in 2002, reading achievement gaps between Native American and white students narrowed at the elementary grade analyzed in three states, according to both the percentages of students scoring proficient and effect sizes. In five states, the Native American-white reading gap narrowed in terms of percentages proficient but widened in terms of effect sizes.

NATIVE AMERICAN STUDENTS, MATH

Trends in gaps between Native American and white students in math are displayed in **table 5-J**. Many states lacked sufficient years of comparable data, or had Native American subgroups too small to analyze. Few states had both effect size and percentage proficient data for this subgroup. At all grade levels analyzed, trends in the Native American-white gap were mixed.

Table 5-J. Gap Trends for Native American Students in Math Since 2002

Type of Trend	Grade	Effect size: Narrowed	Effect size: No change	Effect size: Widened	Effect size: Subgroup small or changed	Effect size: Data unavailable or too few years
Percentage of students scoring at or above the proficient level: Narrowed	E	3 states: AZ, CA, OR	2 states: CO, UT			2 states: FL, OK
	M	2 states: ND, TX		1 state: CO		3 states: FL, MT, OK
	H	1 state: AZ				4 states: AL, CA, MT, RI
Percentage of students scoring at or above the proficient level: No change	E	1 state: NM				1 state: MT
	M	1 state: AZ				
	H					
Percentage of students scoring at or above the proficient level: Widened	E	1 state: WA		3 states: NV, ND, TX		1 state: AK
	M	2 states: NM, WA		1 state: NV		2 states: AL, AK
	H	1 state: WA	1 state: CO	1 state: TX		2 states: NM, OK

Continued on page 87

Table 5-J. Gap Trends for Native American Students in Math Since 2002 (continued)

Type of Trend	Grade	Effect size: →← Narrowed	✪ No change	←→ Widened	✧ Subgroup small or changed	♦/❍ Data unavailable or too few years
Percentage of students scoring at or above the proficient level — ✧ Subgroup small or changed	E				✧ ✧ 9 states: HI, ID, IN, IA, LA, MS, PA, SC, TN	✧ ♦/❍ 8 states: AL, AR, GA, MD, MA, NJ, OH, WV
	M				✧ ✧ 11 states: HI, ID, IN, IA, LA, MS, OR, PA, SC, TN, UT	✧ ♦/❍ 8 states: AR, CA, GA, MD, MA, NJ, OH, WV
	H				✧ ✧ 14 states: HI, CT, ID, IN, IA, LA, MS, NV, ND, OR, PA, SC, TN, UT	✧ ♦/❍ 9 states: AR, FL, GA, ME, MA, NH, NJ, OH, WV
Percentage of students scoring at or above the proficient level — ♦/❍ Data unavailable or too few years	E				♦/❍ ✧ 1 state: DE	♦/❍ ♦/❍ 18 states: CT, IL, KS, KY, ME, MI, MN, MO, NE, NH, NY, NC, RI, SD, VT, VA, WI, WY
	M				♦/❍ ✧ 1 state: DE	♦/❍ ♦/❍ 18 states: CT, IL, KS, KY, ME, MI, MN, MO, NE, NH, NY, NC, RI, SD, VT, VA, WI, WY
	H				♦/❍ ✧ 1 state: DE	♦/❍ ♦/❍ 16 states: AK, IL, KS, KY, MD, MI, MN, MO, NE, NY, NC, SD, VT, VA, WI, WY

Table reads: Since NCLB took effect in 2002, math achievement gaps between Native American and white students narrowed at the elementary grade analyzed in three states, according to both the percentages of students scoring proficient and effect sizes. In two states, the Native American-white math gap narrowed in terms of percentages proficient but showed no change in terms of effect sizes.

SMALL AND CHANGING NATIVE AMERICAN SUBGROUPS

Many states reported small Native American subgroups, and a few states had changing subgroups. For example, 21 states reported having small Native American subgroups at the high school grade tested for NCLB; just 2 states, Oregon and Florida, reported that this subgroup had changed significantly in size at the high school grade analyzed. We analyzed gap trends in states with small or changing Native American enrollments to see whether including these states would have changed the picture of national trends. Again, the results for these small and changing subgroups should be interpreted with caution. In reading and math, gaps narrowed more often than they widened according to both indicators. However, in a number of instances, gaps widened in both subjects according to the percentage proficient indicator.

Table 5-K. Small or Changing Subgroups: Native American-White Gap Trends in Reading

Underlined = Small **Bold** = Changing ***Bold & Italics*** = Both small & changing

Type of Trend		Grade	Effect size: →← Narrowed	Effect size: ✪ No change	Effect size: ←→ Widened	Effect size: ♦/❍ Data unavailable or too few years
Percentage of students scoring at or above the proficient level	Narrowed →←	E	→← →← 7 states: ID, IA, LA, ***MS***, ***PA***, SC, ***TN***			→← ♦/❍ 3 states: MD, OH, WV
		M	→← →← 4 states: ID, IA, LA, ***PA***	→← ✪ 2 states: **OR**, TN	→← ←→ 1 state: ***HI***	→← ♦/❍ 5 states: AR, MD, MA, NJ, **WV**
		H	→← →← 8 states: ***HI***, ID, IA, LA, ND, **OR**, ***PA***, TN	→← ✪ 2 states: ***MS***, NV	→← ←→ 1 state: SC	→← ♦/❍ 6 states: AR, **FL**, ME, MD, MA, NJ
	✪ No change	E				
		M			✪ ←→ 1 state: IN	
		H			✪ ←→ 1 state: ***IN***	

Continued on page 89

Table 5-K. Small or Changing Subgroups: Native American-White Gap Trends in Reading (continued)

Type of Trend	Grade	Effect size: Narrowed (→←)	Effect size: No change (✪)	Effect size: Widened (←→)	Effect size: Data unavailable or too few years (♦/❍)
Percentage at or above the proficient level: Widened (↔)	E		←→ ✪ 1 state: IN	←→ ←→ 1 state: HI	←→ ♦/❍ 4 states: AR, AL, MA, NJ
	M	←→ →← 1 state: ***MS***	←→ ✪ 1 state: SC	←→ ←→ 1 state: **UT**	←→ ♦/❍ 1 state: OH
	H			←→ ←→ 1 state: ***CT***	←→ ♦/❍ 3 states: ***NH, OH,*** WV
Data unavailable or too few years (♦/❍)	E	♦/❍ →← 1 state: DE			
	M			♦/❍ ←→ 1 state: DE	
	H	♦/❍ →← 1 state: DE			

Table reads: Since NCLB took effect in 2002, reading achievement gaps between Native American and white students in states with small or changing subgroups narrowed at the middle school grade analyzed in four states, according to both the percentages of students scoring proficient and effect sizes. In two states, the Native American-white reading gap narrowed at this grade in terms of percentages proficient but showed no change in terms of effect sizes.

Table 5-L. Small or Changing Subgroups: Native American-White Gap Trends in Math

Underlined = Small **Bold** = Changing ***Bold & Italics*** = Both small & changing

Type of Trend	Grade	Effect size: Narrowed (→←)	No change (✪)	Widened (↔)	Data unavailable or too few years (♦/❍)
Percentage of students scoring at or above the proficient level — Narrowed (↓↑)	E	→← →← 5 states: ID, LA, ***MS, PA, TN***	→← ✪ 2 states: IA, SC	→← ↔ 1 state: IN	→← ♦/❍ 5 states: AL, ***GA***, MD, NJ, WV
	M	→← →← 6 states: ID, IA, LA, ***MS***, ***PA***, TN	→← ✪ 1 state: ***OR***	→← ↔ 1 state: **UT**	→← ♦/❍ 2 states: ***GA***, MD
	H	→← →← 6 states: ***HI***, IA, NV, ND, ***PA***, SC		→← ↔ 3 states: ID, LA, ***UT***	→← ♦/❍ 4 states: AR, **FL**, MA, NJ
No change (✪)	E				
	M	✪ →← 1 state: SC			✪ ♦/❍ 2 states: **CA**, MA
	H				
Widened (↕)	E			↔ ↔ 1 state: HI	↔ ♦/❍ 3 states: AR, NJ, OH
	M		↔ ✪ 1 state: IN	↔ ↔ 1 state: ***HI***	↔ ♦/❍ 4 states: AR, NJ, OH, ***WV***
	H	↔ →← 1 state: **OR**		↔ ↔ 4 states: ***CT, IN, MS,*** TN	↔ ♦/❍ 5 states: ***GA***, ME, ***NH, OH,*** WV

Continued on page 91

Table 5-L. Small or Changing Subgroups: Native American-White Gap Trends in Math (continued)

Type of Trend	Grade	Effect size: ➔➔ Narrowed	✪ No change	↔ Widened	♦/○ Data unavailable or too few years
Percentage at or above the proficient level	E	♦/○ ➔➔ 1 state: DE			
♦/○ Data unavailable or too few years	M	♦/○ ➔➔ 1 state: DE			
	H	♦/○ ➔➔ 1 state: DE			

Table reads: Since NCLB took effect in 2002, math achievement gaps between Native American and white students in states with small or changing subgroups narrowed at the elementary grade analyzed in five states, according to both the percentages of students scoring proficient and effect sizes. In two states, the Native American-white math gap narrowed in terms of percentages proficient but showed no change in terms of effect sizes.

STATES WITH THE HIGHEST NATIVE AMERICAN ENROLLMENTS

According to the Common Core of Data (NCES, 2005), in 2005 there were 594,305 American Indian/Alaskan Native students enrolled in public schools. Five states accounted for half (309,948) of the enrollment: Oklahoma, Arizona, California, New Mexico, and Alaska. Similar to the national results shown in tables 5-I and 5-J, gap trends in these five states are mixed and also include many instances of insufficient data:

- Oklahoma did not have effect size data, but gaps in the percentage proficient between Native American and white students narrowed in both reading and math across all grade levels, except for high school math, where gaps widened.
- Results were more mixed for Arizona, California, and New Mexico. Although instances of gaps narrowing outnumbered instances of gaps widening, there were instances in each state where gaps widened or stayed the same.
- Alaska did not have effect size and high school data, but gaps in the percentage proficient widened at the elementary and middle school levels. There were no instances of gaps narrowing.

LOW-INCOME STUDENTS, READING

At the **elementary** level, gaps in reading between students from low-income families and those whose families are not low-income narrowed far more often than they widened (see **table 5-M**). In 10 states, the low-income gaps narrowed in both percentages proficient and effect sizes. In 1 state, the gap widened on both indicators. Of the states with data for only one indicator, gaps shrunk in 12 states and widened in 1. **Middle school** results were also positive. Low-income gaps narrowed in 9 states on both indicators, and in no state did gaps widen. Of the states with percentage proficient data only, gaps narrowed in 11 states and widened in none. At the **high school** level, reading achievement gaps for low-income students narrowed in 4 states on both indicators, and in no state did it widen on both. Of the states with percentage proficient data only, gaps narrowed in 7 states and widened in 2. Across all three grade levels analyzed, there were 7 instances in which gaps narrowed on one indicator and widened on the other.

Table 5-M. Gap Trends for Low-Income Students in Reading Since 2002

Type of Trend		Grade	Effect size: →← Narrowed	✪ No change	←→ Widened	✧ Subgroup small or changed	♦/❍ Data unavailable or too few years
Percentage of students scoring at or above the proficient level	→← Narrowed	E	→← →← 10 states: AZ, HI, ID, IN, IA, KY, NV, NM, NC, PA	→← ✪ 1 state: TX	→← ←→ 3 states: AR, TN, UT[1]		→← ♦/❍ 11 states: AL, AK, FL, LA, MD, MT, NJ, OH, OK, SD, WV
		M	→← →← 9 states: HI, ID, IA, KY, NV, NC, ND, TN, TX		→← ←→ 2 states: AR, UT[1]		→← ♦/❍ 11 states: AL, AK, FL, MD, MT, NJ, OH, OK, SC, SD, WV
		H	→← →← 4 states: AZ, IN, ND, TX	→← ✪ 1 state: ID			→← ♦/❍ 7 states: CA, LA, MD, MT, NM, OK, WV
	✪ No change	E		✪ ✪ 1 state: CA			
		M	✪ →← 1 state: NM	✪ ✪ 1 state: CA			✪ ♦/❍ 1 state: LA
		H					

Continued on page 93

Table 5-M. Gap Trends for Low-Income Students in Reading Since 2002 (continued)

Type of Trend	Grade	Effect size: Narrowed (→←)	Effect size: No change (✪)	Effect size: Widened (←→)	Effect size: Subgroup small or changed (✧)	Effect size: Data unavailable or too few years (♦/○)
Percentage of students scoring at or above the proficient level: Widened (↕)	E			←→ ←→ 1 state: ND		←→ ♦/○ 1 state: SC
	M	←→ →← 1 state: AZ				
	H	←→ →← 1 state: IA				←→ ♦/○ 2 states: AL, SD
Percentage of students scoring at or above the proficient level: Subgroup small or changed (✧)	E				✧ ✧ 1 state: MS	
	M				✧ ✧ 3 states: IN, MS, PA	
	H				✧ ✧ 8 states: CT, HI, KY, MS, PA, SC, TN, UT	✧ ♦/○ 4 states: FL, NH, NJ, OH
Percentage of students scoring at or above the proficient level: Data unavailable or too few years (♦/○)	E	♦/○ →← 1 state: DE	♦/○ ✪ 2 states: CO, OR			♦/○ ♦/○ 18 states: CT, GA, IL, KS, ME, MA, MI, MN, MO, NE, NH, NY, RI, VT, VA, WA, WI, WY
	M	♦/○ →← 1 state: OR	♦/○ ✪ 2 states: CO, DE			♦/○ ♦/○ 18 states: CT, GA, IL, KS, ME, MA, MI, MN, MO, NE, NH, NY, RI, VT, VA, WA, WI, WY
	H		♦/○ ✪ 1 state: OR		♦/○ ✧ 3 states: AR, CO, DE	♦/○ ♦/○ 19 states: AK, GA, IL, KS, ME, MA, MI, MN, MO, NE, NV, NY, NC, RI, VT, VA, WA, WI, WY

Table reads: Since NCLB took effect in 2002, reading achievement gaps between low-income and non-low-income students narrowed at the elementary grade analyzed in 10 states in both percentages proficient and effect sizes. In 3 states, the low-income gap narrowed in terms of percentages proficient but widened in terms of effect sizes.

[1] The percentage proficient data covered a different range of years than the effect size data, which may explain the contradictory trends.

LOW-INCOME STUDENTS, MATH

At the **elementary** level, far more states had evidence of low-income math gaps narrowing than widening, as illustrated in **table 5-N**. In 12 states, gaps shrunk according to both percentages proficient and effect sizes. In no state did gaps widen on both indicators. Of the states with only one type of data, gaps narrowed in 11 states and widened in 2. At the **middle school** level, low-income gaps in math narrowed on both indicators in 8 states and widened on both in 1 state. Of the states with only one type of data, gaps grew smaller in 9 states and larger in 2. Fewer states had sufficient data at the **high school** level, but gaps narrowed on both indicators in 4 states, and in no state did gaps widen on both. Of the states with percentage proficient data only, gaps shrunk in 6 states and widened in 1. Across grades, there were 7 instances of gaps narrowing on one indicator and widening on the other.

Table 5-N. Gap Trends for Low-Income Students in Math Since 2002

Type of Trend		Grade	Effect size: →← Narrowed	✪ No change	←→ Widened	✧ Subgroup small or changed	♦/❍ Data unavailable or too few years
Percentage of students scoring at or above the proficient level	Narrowed	E	→← →← 12 states: AZ, CA, CO, HI, IA, KY, NV, NM, PA, TN, TX, UT	→← ✪ 1 state: ID	→← ←→ 1 state: AR		→← ♦/❍ 9 states: AL, AK, FL, LA, MD, NJ, OH, OK, WV
		M	→← →← 8 states: HI, ID, IA, NV, ND, TN, TX, UT				→← ♦/❍ 7 states: AL, AK, FL, NJ, OH, OK, WV
		H	→← →← 4 states: AZ, IN, ND, TX		→← ←→ 2 states: ID[1], UT[1]		→← ♦/❍ 6 states: AL, CA, LA, MT, NM, WV
	✪ No change	E					
		M	✪ →← 1 state: AZ				✪ ♦/❍ 2 states: LA, MD
		H					

Continued on page 95

Table 5-N. Gap Trends for Low-Income Students in Math Since 2002 (continued)

Type of Trend	Grade	Effect size: Narrowed (→←)	No change (✪)	Widened (←→)	Subgroup small or changed (✧)	Data unavailable or too few years (♦/❍)
Percentage of students scoring at or above the proficient level: Widened (←→)	E	←→ →← 1 state: ND				←→ ♦/❍ 2 states: MT, SC
	M	←→ →← 2 states: KY, NM		←→ ←→ 1 state: AR		←→ ♦/❍ 2 states: MT, SC
	H	←→ →← 1 state: IA				←→ ♦/❍ 1 state: OK
Subgroup small or changed (✧)	E				✧ ✧ 2 states: IN, MS	
	M				✧ ✧ 3 states: IN, MS, PA	✧ ♦/❍ 1 state: CA
	H				✧ ✧ 7 states: CT, HI, KY, MS, PA, SC, TN	✧ ♦/❍ 4 states: FL, NH, NJ, OH
Data unavailable or too few years (♦/❍)	E	♦/❍ →← 2 states: DE, OR				♦/❍ ♦/❍ 20 states: CT, GA, IL, KS, ME, MA, MI, MN, MO, NE, NH, NY, NC, RI, SD, VT, VA, WA, WI, WY
	M	♦/❍ →← 2 states: CO, OR			♦/❍ ✧ 1 state: DE	♦/❍ ♦/❍ 20 states: CT, GA, IL, KS, ME, MA, MI, MN, MO, NE, NH, NY, NC, RI, SD, VT, VA, WA, WI, WY
	H			♦/❍ ←→ 1 state: OR	♦/❍ ✧ 3 states: AR, CO, DE	♦/❍ ♦/❍ 21 states: AK, GA, IL, KS, ME, MD, MA, MI, MN, MO, NE, NV, NY, NC, RI, SD, VT, VA, WA, WI, WY

Table reads: Since NCLB took effect in 2002, math achievement gaps between low-income and non-low-income students narrowed at the elementary grade analyzed in 12 states on both percentages proficient and effect sizes. In 1 state, the low-income math gap narrowed in terms of percentages proficient but widened in terms of effect sizes.

[1] The percentage proficient data covered a different range of years than the effect size data, which may explain the contradictory trends.

SMALL AND CHANGING LOW-INCOME SUBGROUPS

In all of the states listed in the "subgroup small or changed" columns of the previous two tables, the low-income subgroup changed in size by more than 25%. No state was listed because this subgroup was too small. The low-income subgroup was excluded from the initial analysis in fewer states than other subgroups were. In the states excluded from the main analyses, the low-income achievement gap narrowed more often than it widened in both reading and math.

STATES WITH THE HIGHEST NUMBERS OF CHILDREN IN POVERTY

The Common Core of Data (NCES, 2005) has no "low-income" category consisting of students eligible for free or reduced-price lunches—the main criteria used to determine low-income students for NCLB accountability—so we substituted data from the 2000 census for children aged 5-17 in poverty (U.S. Census Bureau, n.d.). That year, the U.S. had 8,334,825 children in poverty. The eight states that accounted for more than half (4,386,105) of all 5- to 17-year-olds in poverty were California, Texas, New York, Florida, Illinois, Pennsylvania, Ohio, and Georgia. Because New York, Illinois, and Georgia lacked data for this analysis, we focused on the remaining five.

- The gap picture for low-income students in these states is positive. There were no instances of any gaps growing wider at any grade level. Where data were available, gaps mostly narrowed. They showed no net change in a few instances, namely reading at the elementary and middle school levels in California.

- Florida and Ohio lacked effect size data, but gaps narrowed at all levels in terms of percentages proficient. Gaps also narrowed in Pennsylvania on both measures, but results for the upper grades should be viewed with caution because of changes in the size of the low-income subgroup.

Possible Explanations

Taken together, our gap analyses show many more states with evidence of achievement gaps narrowing than widening since NCLB was enacted in 2002. Three of the four subgroups analyzed—African American, Latino, and low-income students—had many more instances of gaps narrowing than widening. The exception was the Native American subgroup, where results were more mixed and the analysis was hampered by lack of data. Even though gaps have narrowed, they remain wide for major subgroups in many states. The magnitude of the gaps, reported in more detail in chapter 6, suggests that closing achievement gaps is a long-term proposition that will require continued and intensive efforts.

Below we highlight some possible explanations for our finding that achievement gaps on state tests have narrowed since 2002 much more often than they have widened. Some of these explanations overlap with those given for improvements in overall achievement at the end of chapter 3. Any or all of these factors in combination may be contributing to achievement gap trends, and different explanations could apply to different states, districts, or schools:

- **Better instruction for subgroups.** The most hopeful explanation of narrowing gaps is that students in the targeted subgroups are actually learning more. NCLB places immense pressure on states, schools, and districts to improve the achievement of lower-performing subgroups. CEP's past studies have amassed evidence that educators are focusing more attention on subgroups; they have instituted special programs or have devoted more resources to rais-

ing achievement for lower-performing subgroups. CEP case studies suggest that students from underperforming subgroups are receiving more intensive instruction and extended instructional time than in the past and seem to be learning more (CEP, 2006).

- **More familiarity with tests.** Educators may be spending more time with students in subgroups to better prepare them for tests, familiarizing them with test instructions and question formats and what to expect. This can be appropriate as long as it is limited and most of the instruction still focuses on the state content standards.

- **Score inflation.** A less optimistic explanation is score inflation. Targeted subgroups may be spending most of their classroom time focusing on the content that is likely to show up on the tests, rather than on learning the broader domains of reading and math. This kind of inappropriate teaching to the test can make test scores go up without real learning gains.

- **Urban areas.** Student subgroups are concentrated in urban areas, and urban districts are those most affected by NCLB. For school year 2005-06, 43% of urban districts reported having at least one school identified for improvement under NCLB, compared with 14% of suburban districts and 11% of rural districts (CEP, 2006). Therefore, urban districts may feel a great deal more pressure to boost test scores than their suburban or rural counterparts and may be more likely to reallocate resources toward this end. Other studies have also shown progress by large urban school districts in narrowing achievement gaps (Snipes et al., 2008).

- **Small manipulations affecting distribution of scores.** This study takes into account publicly announced changes in testing programs or cut scores, but as mentioned in our possible explanation of overall achievement trends in chapter 3, sometimes small manipulations are made that could make a test easier but are not entirely transparent. Overall, we found that the percentage proficient indicator gives a more positive view of progress on achievement gaps than the effect size indicator. This lends support to the possibility that in some states, small manipulations in testing policies could be affecting the distribution of test scores relative to the proficiency cut score, which could make gaps appear smaller according to the percentage proficient measure.

- **Decrease in white student achievement.** Another possible explanation is that achievement gaps have narrowed because the achievement of white students has decreased on the whole since 2002. As a quick check, we analyzed achievement trends using the percentage proficient for white students in all 50 states to see how many states posted declines for this group. We found that this explanation does not hold for the elementary and middle school levels. At both of these grade levels, achievement for the white subgroup went up in a majority of states. At the elementary level, the performance of the white subgroup increased in reading in 32 of the 33 states with sufficient data, and increased in math in 31 of 32 states. At the middle school level, white performance increased in reading in 27 of 33 states with sufficient data, and in math in 30 of 32 states. The high school results, however, were more mixed. In reading, white high school student achievement increased in 22 states but declined in 10. In math, achievement by this subgroup increased in 26 states but declined in 7. Therefore, it is possible that at the high school level, particularly in reading, the narrowing of achievement gaps in some states can be attributed to declines in the percentage proficient for the white subgroup. A more detailed analysis would have to be performed to determine the extent to which this was a factor at the high school level.

Table 5-O. State-by-State Summary of Achievement Gap Trends in Reading

Legend

PP = Percentage proficient

ES = Effect size

→← = Gap narrowed

←→ = Gap widened

✱ = Mixed (gaps showed different trends at different grade levels)

✪ = No change in gap

♦ = Not enough years of data (only 1-2 years) to determine trend

❍ = Data not available

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
Alabama *(2004–2007)*								
Across all grade levels	✱		→←		✱[2]		✱	
Elementary (grade 4)	→←	♦	→←	♦	←→ [2]	♦	→←	♦
Middle (grade 8)	→←	♦	→←	♦	→←	♦	→←	♦
High (grade 11)	←→	♦	→←	♦	←→	♦	←→	♦
Alaska *(2005–2007)*								
Across all grade levels								
Elementary (grade 4)	→← [2]	❍	←→ [2]	❍	←→	❍	→←	❍
Middle (grade 8)	→← [2]	❍	→← [2]	❍	←→	❍	→←	❍
High (grade 10)	♦	❍	♦	❍	♦	❍	♦	❍
Arizona *(2005–2007)*								
Across all grade levels	✱[3]	→← [3]	✱[3]	←→ [3]	✱	→←	✱	→←
Elementary (grade 4)	→←	→←	→←	→←	→←	→←	→←	→←
Middle (grade 8)	←→	→←	✪	→←	←→	→←	←→	→←
High (grade 10)	←→ [3]	→← [3]	→← [3]	→← [3]	✪	→←	→←	→←
Arkansas *(2002–2007)*								
Across all grade levels	✱	←→						←→ [3]
Elementary (grade 4)	✪	←→	←→ [3]	←→ [3]	←→ [2]	♦	→←	←→
Middle (grade 8)	→←	←→	←→ [3]	←→ [3]	→← [2]	♦	→←	←→
High (grade 11)	←→	←→	→← [3]	→← [3]	→← [2]	♦	♦	←→ [3]
California *(2003–2007)*								
Across all grade levels	✱		✱		✱		✱[1]	
Elementary (grade 4)	✪	✪	→←	→←	→←	←→	✪[1]	✪[1]
Middle (grade 8)	←→	✪	←→	←→	←→	✪	✪[1]	✪[1]
High (grade 10)	→←	❍	✪	❍	→←	❍	→← [1]	❍

Continued on page 99

Table 5-O. State-by-State Summary of Achievement Gap Trends in Reading (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
Colorado *(2002–2007)*								
Across all grade levels	✱	✱	→← 3	✱ 3	✱	✱		✱ 3
Elementary (grade 4)	→←	→←	→←	✪	←→	→←	♦	✪
Middle (grade 8)	←→	←→	→←	✪	✪	←→	♦	✪
High (grade 10)	←→	←→	→← 3	→← 3	→←	←→	♦	←→ 3
Connecticut *(2002–2006)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 10)	→← 3	←→ 3	→← 3	←→ 3	←→ 2, 3	→← 2, 3	→← 3	←→ 3
Delaware *(2002–2007)*								
Across all grade levels		✱ 3						✱ 3
Elementary (grade 5)	♦	→←	♦	→← 3	♦	→← 2	♦	→←
Middle (grade 8)	♦	←→ 3	♦	✪ 3	♦	←→ 2	♦	✪
High (grade 10)	♦	→←	♦	→← 3	♦	→← 2	♦	→← 3
Florida *(2002–2007)*								
Across all grade levels	→← 3	✱ 3	→← 3	→← 3	✱ 3		→← 3	
Elementary (grade 4)	→←	✪	→←	→←	←→	❍	→←	❍
Middle (grade 8)	→←	→←	→← 3	→← 3	←→	❍	→←	❍
High (grade 10)	→← 3	→← 3	→← 3	→← 3	→← 3	❍	→← 3	❍
Georgia *(2002–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 11)	♦	❍	♦	❍	♦	❍	♦	❍
Hawaii *(2002–2006)*								
Across all grade levels							✱ 3	✱ 3
Elementary (grade 5)	→← 2	→← 2	→← 2, 3	→← 2, 3	←→ 2	←→ 2	→←	→←
Middle (grade 8)	←→ 2	✪ 2	→← 2, 3	→← 2, 3	→← 2, 3	←→ 2, 3	→←	→←
High (grade 10)	→← 2, 3	→← 2, 3	→← 2, 3	→← 2, 3	→← 2, 3	→← 2, 3	←→ 3	✪ 3
Idaho *(2003–2007)*								
Across all grade levels			→← 3	✱ 3			→← 1	✱ 1
Elementary (grade 4)	←→ 2	←→ 2	→←	→←	→← 2	→← 2	→← 1	→← 1
Middle (grade 8)	→← 2	←→ 2	→←	←→	→← 2	→← 2	→← 1	→← 1
High (grade 10)	→← 2, 3	→← 2, 3	→← 3	→← 3	→← 2	→← 2	→← 1	✪ 1

Continued on page 100

Table 5-O. State-by-State Summary of Achievement Gap Trends in Reading (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
Illinois *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 11)	♦	♦	♦	♦	♦	♦	♦	♦
Indiana *(2002–2007)*								
Across all grade levels	✱	✱					→← 3	→← 3
Elementary (grade 4)	→←	→←	←→ 3	←→ 3	←→ 2	✪ 2	→←	→←
Middle (grade 8)	→←	→←	←→ 3	←→ 3	✪ 2	←→ 2	→← 3	→← 3
High (grade 10)	✪	←→	→← 3	→← 3	✪ 2, 3	←→ 2, 3	→←	→←
Iowa *(2004–2007)*								
Across all grade levels	→←	→←	→← 3	→← 3			✱ 1	→← 1
Elementary (grade 4)	→←	→←	→←	→←	→← 2	→← 2	→← 1	→← 1
Middle (grade 8)	→←	→←	→←	→←	→← 2	→← 2	→← 1	→← 1
High (grade 11)	→←	→←	→← 3	→← 3	→← 2	→← 2	←→ 1	→← 1
Kansas *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 11)	♦	♦	♦	♦	♦	♦	♦	♦
Kentucky *(2002–2006)*								
Across all grade levels	✱	→←					✱ 3	→← 3
Elementary (grade 4)	→←	→←	✪ 3	✪ 3	❍	❍	→←	→←
Middle (grade 7)	→←	→←	←→ 3	←→ 3	❍	❍	→←	→←
High (grade 10)	←→	→←	←→ 3	←→ 3	❍	❍	←→ 3	→← 3
Louisiana *(2002–2007)*								
Across all grade levels	→←	→←	✱ 3	←→ 3			✱	
Elementary (grade 4)	→←	→←	←→ 3	←→ 3	→← 2	→← 2	→←	❍
Middle (grade 8)	→←	→←	←→ 3	←→ 3	→← 2	→← 2	✪	❍
High (grade 10)	→←	→←	→←	←→	→← 2	→← 2	→←	❍
Maine *(2003–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 11)	→← 2, 3	♦	→← 2	♦	→← 2	♦	♦	♦

Continued on page 101

Table 5-O. State-by-State Summary of Achievement Gap Trends in Reading (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
Maryland *(2003–2007)*								
Across all grade levels	→←		✱[3]				→←	
Elementary (grade 4)	→←	❍	→←	❍	→←[2]	❍	→←	❍
Middle (grade 8)	→←	❍	→←[3]	❍	→←[2]	❍	→←	❍
High (Eng II EOC)	→←	❍	←→[3]	❍	→←[2]	❍	→←	❍
Massachusetts *(2002–2007)*								
Across all grade levels	✱		→←					
Elementary (grade 4)	→←	❍	→←	❍	←→[2]	❍	❍	❍
Middle (grade 7)	←→	❍	→←	❍	→←[2]	❍	❍	❍
High (grade 10)	→←	❍	→←	❍	→←[2]	❍	❍	❍
Michigan *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 11)	♦	♦	♦	♦	♦	♦	♦	♦
Minnesota *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 10)	♦	♦	♦	♦	♦	♦	♦	♦
Mississippi *(2002–2007)*								
Across all grade levels	→←	→←						
Elementary (grade 4)	→←	→←	→←[3]	←→[3]	→←[2, 3]	→←[2, 3]	→←[3]	←→[3]
Middle (grade 8)	→←	→←	←→[2, 3]	←→[2, 3]	←→[2, 3]	→←[2, 3]	←→[3]	←→[3]
High (grade varies)	→←	→←	→←[2, 3]	✪[2, 3]	→←[2, 3]	✪[2, 3]	→←[3]	←→[3]
Missouri *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 11)	♦	♦	♦	♦	♦	♦	♦	♦
Montana *(2004–2007)*								
Across all grade levels					→←		→←	
Elementary (grade 4)	→←[2]	♦	→←[2]	♦	→←	♦	→←	♦
Middle (grade 8)	→←[2]	♦	→←[2]	♦	→←	♦	→←	♦
High (grade 10)	→←[2, 3]	♦	→←[2]	♦	→←	♦	→←	♦

Continued on page 102

Table 5-O. State-by-State Summary of Achievement Gap Trends in Reading (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
Nebraska *(2002–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	❍	♦	❍	♦	❍	♦	❍
Middle (grade 8)	♦	❍	♦	❍	♦	❍	♦	❍
High (grade 11)	♦	❍	♦	❍	♦	❍	♦	❍
Nevada *(2004–2007)*								
Across all grade levels	→← [3]	✱ [3]	→← [3]	→← [3]	✱ [2]	✱ [2]		
Elementary (grade 5)	→←	→←	→←	→←	→←	←→	→←	→←
Middle (grade 8)	→←	→←	→← [3]	→← [3]	←→	←→	→←	→←
High (grade 10)	→← [3]	✪ [3]	→← [3]	→← [3]	→← [2]	✪ [2]	♦	♦
New Hampshire *(2002–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	❍	♦	❍	♦	❍	♦	❍
Middle (grade 8)	♦	❍	♦	❍	♦	❍	♦	❍
High (grade 10)	←→ [2, 3]	❍	←→ [2, 3]	❍	←→ [2, 3]	❍	←→ [3]	❍
New Jersey *(2002–2007)*								
Across all grade levels	→←		→← [3]				→← [1, 3]	
Elementary (grade 4)	→←	❍	→←	❍	←→ [2]	❍	→← [1]	❍
Middle (grade 8)	→←	❍	→← [3]	❍	→← [2]	❍	→← [1]	❍
High (grade 11)	→←	❍	→←	❍	→← [2]	❍	→← [1, 3]	❍
New Mexico *(2005–2007)*								
Across all grade levels	✱ [2]		✱		✱		✱	
Elementary (grade 4)	→←	→←	→←	→←	→←	→←	→←	→←
Middle (grade 8)	→←	←→	→←	→←	←→	→←	✪	→←
High (grade 11)	←→ [2]	♦	←→	♦	←→	♦	→←	♦
New York *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	❍	♦	❍	♦	❍	♦	❍
Middle (grade 8)	♦	❍	♦	❍	♦	❍	♦	❍
High (grade varies)	❍	❍	❍	❍	❍	❍	❍	❍
North Carolina *(2003–2007)*								
Across all grade levels								
Elementary (grade 4)	→←	←→	→← [3]	✪ [3]	→←	←→	→←	→←
Middle (grade 8)	→←	✪	→← [3]	→← [3]	→←	←→	→←	→←
High (grade 10)	❍	❍	❍	❍	❍	❍	❍	❍

Continued on page 103

Table 5-O. State-by-State Summary of Achievement Gap Trends in Reading (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
North Dakota *(2005–2007)*								
Across all grade levels					→← 2	✱ 2	✱	✱
Elementary (grade 4)	→← 2, 3	←→ 2, 3	→← 2, 3	←→ 2, 3	→←	✪	←→	←→
Middle (grade 8)	←→ 2	←→ 2	→← 2	←→ 2	→←	→←	→←	→←
High (grade 11)	←→ 2, 3	←→ 2, 3	→← 2	→← 2	→← 2	→← 2	→←	→←
Ohio *(2004–2007)*								
Across all grade levels	✱ 3		→← 3				→← 3	
Elementary (grade 4)	←→	❍	→←	❍	→← 2	❍	→←	❍
Middle (grade 8)	→←	❍	→←	❍	←→ 2	❍	→←	❍
High (OH Grad Test)	→← 3	❍	→← 3	❍	←→ 2, 3	❍	→← 3	❍
Oklahoma *(2002–2007)*								
Across all grade levels	→←		✱		→←		→← 1	
Elementary (grade 5)	→←	❍	→←	❍	→←	❍	→← 1	❍
Middle (grade 8)	→←	❍	→←	❍	→←	❍	→← 1	❍
High (English II)	→←	❍	✪	❍	→←	❍	→← 1	❍
Oregon *(2002–2007)*								
Across all grade levels	✱ 3	←→ 3			→← 3	✱ 3		✱ 1
Elementary (grade 5)	→←	←→	→← 3	←→ 3	→←	→←	♦	✪ 1
Middle (grade 8)	→←	←→	←→ 3	←→ 3	→← 3	✪ 3	♦	→← 1
High (grade 10)	←→ 3	←→ 3	→← 3	←→ 3	→← 3	→← 3	♦	✪ 1
Pennsylvania *(2002–2007)*								
Across all grade levels	→← 3	→← 3					→← 1, 3	→← 1, 3
Elementary (grade 5)	→←	→←	→← 3	→← 3	→← 2, 3	→← 2, 3	→← 1	→← 1
Middle (grade 8)	→← 3	→← 3	→← 3	→← 3	→← 2, 3	→← 2, 3	→← 1, 3	→← 1, 3
High (grade 11)	→← 3	→← 3	✪ 2, 3	✪ 2, 3	→← 2, 3	→← 2, 3	→← 1, 3	→← 1, 3
Rhode Island *(2004–2006)*								
Across all grade levels								
Elementary (grade 4)	♦	❍	♦	❍	♦	❍	♦	❍
Middle (grade 8)	♦	❍	♦	❍	♦	❍	♦	❍
High (grade 11)	←→	❍	←→	❍	→←	❍	❍	❍
South Carolina *(2002–2007)*								
Across all grade levels	✱	✱					✱ 1, 3	
Elementary (grade 4)	✪	→←	←→ 3	✪ 3	→← 2	→← 2	←→ 1	❍
Middle (grade 8)	→←	→←	←→ 3	←→ 3	←→ 2	✪ 2	→← 1	❍
High (grade 10)	→←	✪	→← 3	←→ 3	→← 2	←→ 2	→← 1, 3	→← 1, 3

Continued on page 104

Table 5-O. State-by-State Summary of Achievement Gap Trends in Reading (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
South Dakota *(2005–2007)*								
Across all grade levels					✱		✱[1]	
Elementary (grade 4)	↔[2]	○	↔[2]	○	→←	○	→←[1]	○
Middle (grade 8)	↔[2]	○	→←[2]	○	↔	○	→←[1]	○
High (grade 11)	↔[2]	○	→←[2]	○	↔	○	↔[1]	○
Tennessee *(2004–2007)*								
Across all grade levels	→←	✱					→←[3]	✱[3]
Elementary (grade 4)	→←	↔	→←[3]	↔[3]	→←[2, 3]	→←[2, 3]	→←	↔
Middle (grade 8)	→←	→←	→←[3]	→←[3]	→←[2]	✪[2]	→←	→←
High (grade 10)	→←	→←	→←[3]	↔[3]	→←[2]	→←[2]	→←[3]	→←[3]
Texas *(2005–2007)*								
Across all grade levels	→←	→←	→←	✱	✱	✱	→←	✱
Elementary (grade 4)	→←	→←	→←	↔	↔	↔	→←	✪
Middle (grade 8)	→←	→←	→←	→←	→←	→←	→←	→←
High (grade 10)	→←	→←	→←	→←	→←	↔	→←	→←
Utah *(2004–2007)*								
Across all grade levels	✱[2]	✱[2]	→←	↔	✱[3]	↔[3]	→←[3]	✱[3]
Elementary (grade 4)	↔	↔	→←	↔	→←	↔	→←	↔
Middle (grade 8)	→←[2]	→←[2]	→←	↔	↔[3]	↔[3]	→←	↔
High (grade 10)	↔[2]	↔[2]	→←	↔	→←	↔	→←[3]	→←[3]
Vermont *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	○	♦	○	♦	○	♦	○
Middle (grade 8)	♦	○	♦	○	♦	○	♦	○
High (grade 10)	♦	○	♦	○	♦	○	♦	○
Virginia *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	○	♦	○	♦	○	♦	○
Middle (grade 8)	♦	○	♦	○	♦	○	♦	○
High (grade 11)	♦	○	♦	○	♦	○	♦	○
Washington *(2002–2007)*								
Across all grade levels	→←[3]	→←[3]			→←	✱		
Elementary (grade 4)	→←	→←	→←[3]	→←[3]	→←	↔	♦	♦
Middle (grade 7)	→←	→←	→←[3]	→←[3]	→←	→←	♦	♦
High (grade 10)	→←[3]	→←[3]	→←[3]	→←[3]	→←	↔	♦	♦

Continued on page 105

Table 5-O. State-by-State Summary of Achievement Gap Trends in Reading (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
West Virginia *(2004–2007)*								
Across all grade levels	→←	→←					→←	
Elementary (grade 4)	→←	→←	→← [2, 3]	❍	→← [2]	❍	→←	❍
Middle (grade 8)	→←	→←	←→ [2, 3]	❍	→← [2, 3]	❍	→←	❍
High (grade 10)	→←	→←	←→ [2]	❍	←→ [2]	❍	→←	❍
Wisconsin *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 10)	♦	♦	♦	♦	♦	♦	♦	♦
Wyoming *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 11)	♦	♦	♦	♦	♦	♦	♦	♦

Table reads: In Alabama, the gap between African American and white students narrowed at the elementary and middle grades analyzed for this report according to the percentage of students scoring at the proficient level or above on the state reading tests. At the high school level, the African American-white gap widened. There were not enough years (only 1-2 years) of effect size data to determine the gap trend according to this measure.

Note: Gap trends for students with disabilities and English language learners are not shown because state and federal policy changes may have affected the year-to-year comparability of test results for these subgroups. Trends for Asian students are not shown because in most states this subgroup performed as well as or better than white students.

[1] Low-income students were compared with all tested students in the state rather than with students who are not low-income.

[2] The number of students tested in this subgroup at this grade level was fewer than 500, so changes for this subgroup should be interpreted with caution.

[3] The number of students tested in this subgroup changed by at least +/-25% between the first and last year shown, so changes in results over time may be due to changes in the composition of the subgroup as well as changes in achievement.

Table 5-P. State-by-State Summary of Achievement Gap Trends in Math

Legend

PP = Percentage proficient

ES = Effect size

→← = Gap narrowed

←→ = Gap widened

✱ = Mixed (gaps showed different trends at different grade levels)

✪ = No change in gap

♦ = Not enough years of data (only 1-2 years) to determine trend

❍ = Data not available

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
Alabama *(2004–2007)*								
Across all grade levels	→←		✱		✱[2]		→←	
Elementary (grade 4)	→←	♦	→←	♦	→←[2]	♦	→←	♦
Middle (grade 8)	→←	♦	←→	♦	←→	♦	→←	♦
High (grade 11)	→←	♦	→←	♦	→←	♦	→←	♦
Alaska *(2005–2007)*								
Across all grade levels								
Elementary (grade 4)	→←[2]	❍	→←[2]	❍	←→	❍	→←	❍
Middle (grade 8)	→←[2]	❍	→←[2]	❍	←→	❍	→←	❍
High (grade 10)	♦	❍	♦	❍	♦	❍	♦	❍
Arizona *(2005–2007)*								
Across all grade levels	✱[3]	→←[3]	→←[3]	→←[3]	✱	→←	✱	→←
Elementary (grade 4)	→←	→←	→←	→←	→←	→←	→←	→←
Middle (grade 8)	→←	→←	→←	→←	✪	→←	✪	→←
High (grade 10)	✪[3]	→←[3]	→←[3]	→←[3]	→←	→←	→←	→←
Arkansas *(2002–2007)*								
Across all grade levels	←→[3]	✱[3]						←→[3]
Elementary (grade 4)	←→	✪	←→[3]	←→[3]	←→[2]	♦	→←	←→
Middle (grade 8)	←→	←→	→←[3]	←→[3]	←→[2]	♦	←→	←→
High (Algebra I)	←→[3]	→←[3]	←→[3]	✪[3]	→←[2]	♦	♦	←→[3]
California *(2003–2007)*								
Across all grade levels	✱[3]		→←[3]		✱[3]		→←[1, 3]	
Elementary (grade 4)	→←	→←	→←	→←	→←	→←	→←[1]	→←[1]
Middle (Algebra I)	✪[3]	❍	→←[3]	❍	✪[3]	❍	→←[1, 3]	❍
High (grade 10)	→←	❍	→←	❍	→←	❍	→←[1]	❍

Continued on page 107

Table 5-P. State-by-State Summary of Achievement Gap Trends in Math (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
Colorado *(2002–2007)*								
Across all grade levels	→←	✱	→←[3]	✱[3]	✱	✱		✱[3]
Elementary (grade 4)	→←	→←	→←	✪	→←	✪	→←	→←
Middle (grade 8)	→←	←→	→←	→←	→←	←→	♦	→←
High (grade 10)	→←	→←	→←[3]	✪[3]	←→	✪	♦	←→[3]
Connecticut *(2002–2006)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 10)	←→[3]	←→[3]	→←[3]	→←[3]	←→[2, 3]	←→[2, 3]	→←[3]	✪[3]
Delaware *(2002–2007)*								
Across all grade levels		✱[3]						→←[3]
Elementary (grade 5)	♦	→←	♦	→←[3]	♦	→←[2]	♦	→←
Middle (grade 8)	♦	✪[3]	♦	→←[3]	♦	→←[2]	♦	→←[3]
High (grade 10)	♦	→←	♦	→←[3]	♦	→←[2]	♦	→←[3]
Florida *(2002–2007)*								
Across all grade levels	→←[3]	→←[3]	→←[3]	→←[3]	→←[3]		→←[3]	
Elementary (grade 4)	→←	→←	→←	→←	→←	❍	→←	❍
Middle (grade 8)	→←	→←	→←[3]	→←[3]	→←	❍	→←	❍
High (grade 10)	→←[3]	→←[3]	→←[3]	→←[3]	→←[3]	❍	→←[3]	❍
Georgia *(2002–2007)*								
Across all grade levels	→←[3]							
Elementary (grade 4)	→←	♦	→←[3]	♦	→←[2, 3]	♦	❍	♦
Middle (grade 8)	→←[3]	♦	→←	♦	→←[2, 3]	♦	❍	♦
High (grade 11)	→←[3]	❍	→←[3]	❍	←→[2, 3]	❍	❍	❍
Hawaii *(2002–2006)*								
Across all grade levels							→←[3]	→←[3]
Elementary (grade 5)	→←[2]	→←[2]	→←[2]	→←[2]	←→[2]	←→[2]	→←	→←
Middle (grade 8)	→←[2]	→←[2]	→←[2, 3]	→←[2, 3]	←→[2, 3]	←→[2, 3]	→←	→←
High (grade 10)	→←[2, 3]	→←[2, 3]	→←[2, 3]	→←[2, 3]	→←[2, 3]	→←[2, 3]	→←[3]	→←[3]
Idaho *(2003–2007)*								
Across all grade levels			→←[3]	✱[3]			→←[1]	✱[1]
Elementary (grade 4)	→←[2]	←→[2]	→←	←→	→←[2]	→←[2]	→←[1]	✪[1]
Middle (grade 8)	→←[2]	←→[2]	→←	→←	→←[2]	→←[2]	→←[1]	→←[1]
High (grade 10)	→←[2, 3]	→←[2, 3]	→←[3]	→←[3]	→←[2]	←→[2]	→←[1]	←→[1]

Continued on page 108

Table 5-P. State-by-State Summary of Achievement Gap Trends in Math (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
Illinois *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 11)	♦	♦	♦	♦	♦	♦	♦	♦
Indiana *(2002–2007)*								
Across all grade levels	→←	→←					→←[3]	→←[3]
Elementary (grade 4)	→←	→←	✪[3]	✪[3]	→←[2]	←→[2]	→←[3]	→←[3]
Middle (grade 8)	→←	→←	←→[3]	←→[3]	←→[2]	✪[2]	→←[3]	→←[3]
High (grade 10)	→←	→←	→←[3]	→←[3]	←→[2, 3]	←→[2, 3]	→←	→←
Iowa *(2004–2007)*								
Across all grade levels	→←	→←	→←[3]	→←[3]			✱[1]	→←[1]
Elementary (grade 4)	→←	→←	→←	→←	→←[2]	✪[2]	→←[1]	→←[1]
Middle (grade 8)	→←	→←	→←	→←	→←[2]	→←[2]	→←[1]	→←[1]
High (grade 11)	→←	→←	→←[3]	→←[3]	→←[2]	→←[2]	←→[1]	→←[1]
Kansas *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 10)	♦	♦	♦	♦	♦	♦	♦	♦
Kentucky *(2002–2006)*								
Across all grade levels	✱[3]	✱[3]					✱[3]	→←[3]
Elementary (grade 5)	→←	→←	→←[3]	←→[3]	❍	❍	→←	→←
Middle (grade 8)	←→	✪	←→[3]	←→[3]	❍	❍	←→	→←
High (grade 11)	←→[3]	→←[3]	←→[3]	←→[3]	❍	❍	✪[3]	→←[3]
Louisiana *(2002–2007)*								
Across all grade levels	→←	→←	✱[3]	✱[3]			✱	
Elementary (grade 4)	→←	→←	←→[3]	←→[3]	→←[2]	→←[2]	→←	❍
Middle (grade 8)	→←	→←	←→[3]	←→[3]	→←[2]	→←[2]	✪	❍
High (grade 10)	→←	→←	→←	✪	→←[2]	←→[2]	→←	❍
Maine *(2003–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 11)	←→[2, 3]	♦	←→[2]	♦	←→[2]	♦	♦	♦

Continued on page 109

Table 5-P. State-by-State Summary of Achievement Gap Trends in Math (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
Maryland *(2003–2007)*								
Across all grade levels								
Elementary (grade 4)	→←	❍	→←	❍	→←[2]	❍	→←	❍
Middle (grade 8)	←→	❍	←→[3]	❍	→←[2]	❍	✪	❍
High (Algebra)	♦	❍	♦	❍	♦	❍	♦	❍
Massachusetts *(2002–2007)*								
Across all grade levels	✱		✱					
Elementary (grade 4)	→←	❍	→←	❍	←→[2]	❍	❍	❍
Middle (grade 8)	←→	❍	←→	❍	✪[2]	❍	❍	❍
High (grade 10)	→←	❍	→←	❍	→←[2]	❍	❍	❍
Michigan *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 11)	♦	♦	♦	♦	♦	♦	♦	♦
Minnesota *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 11)	♦	♦	♦	♦	♦	♦	♦	♦
Mississippi *(2002–2007)*								
Across all grade levels	→←	→←						
Elementary (grade 4)	→←	→←	←→[3]	→←[3]	→←[2, 3]	→←[2, 3]	→←[3]	←→[3]
Middle (grade 8)	→←	→←	←→[2, 3]	←→[2, 3]	→←[2, 3]	→←[2, 3]	→←[3]	←→[3]
High (grade varies)	→←	→←	✪[2, 3]	←→[2, 3]	←→[2, 3]	←→[2, 3]	→←[3]	←→[3]
Missouri *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 10)	♦	♦	♦	♦	♦	♦	♦	♦
Montana *(2004–2007)*								
Across all grade levels					✱		✱	
Elementary (grade 4)	→←[2]	♦	✪[2]	♦	✪	♦	←→	♦
Middle (grade 8)	→←[2]	♦	→←[2]	♦	→←	♦	←→	♦
High (grade 10)	→←[2, 3]	♦	→←[2]	♦	→←	♦	→←	♦

Continued on page 110

Table 5-P. State-by-State Summary of Achievement Gap Trends in Math (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
Nebraska *(2002–2007)*								
Across all grade levels								
Elementary (grade 4)	❍	❍	❍	❍	❍	❍	❍	❍
Middle (grade 8)	❍	❍	❍	❍	❍	❍	❍	❍
High (grade 11)	❍	❍	❍	❍	❍	❍	❍	❍
Nevada *(2004–2007)*								
Across all grade levels	✱[3]	✱[3]	→← [3]	→← [3]	✱[2]	✱[2]		
Elementary (grade 5)	✪	→←	→←	→←	←→	←→	→←	→←
Middle (grade 8)	→←	✪	→← [3]	→← [3]	←→	←→	→←	→←
High (grade 10)	←→ [3]	→← [3]	→← [3]	→← [3]	→← [2]	→← [2]	♦	♦
New Hampshire *(2002–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	❍	♦	❍	♦	❍	♦	❍
Middle (grade 8)	♦	❍	♦	❍	♦	❍	♦	❍
High (grade 10)	←→ [2, 3]	❍	←→ [2, 3]	❍	←→ [2, 3]	❍	←→ [3]	❍
New Jersey *(2002–2007)*								
Across all grade levels	→←		→← [3]				→← [1, 3]	
Elementary (grade 4)	→←	❍	→←	❍	→← [2]	❍	→← [1]	❍
Middle (grade 8)	→←	❍	→← [3]	❍	←→ [2]	❍	→← [1]	❍
High (grade 11)	→←	❍	→←	❍	→← [2]	❍	→← [1, 3]	❍
New Mexico *(2005–2007)*								
Across all grade levels	✱[2]		✱		✱		✱	
Elementary (grade 4)	→←	→←	→←	→←	✪	→←	→←	→←
Middle (grade 8)	←→	→←	←→	→←	←→	→←	←→	→←
High (grade 11)	→← [2]	♦	→←	♦	←→	♦	→←	♦
New York *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	❍	♦	❍	♦	❍	♦	❍
Middle (grade 8)	♦	❍	♦	❍	♦	❍	♦	❍
High (grade varies)	❍	❍	❍	❍	❍	❍	❍	❍
North Carolina *(2003–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 10)	❍	❍	❍	❍	❍	❍	❍	❍

Continued on page 111

Table 5-P. State-by-State Summary of Achievement Gap Trends in Math (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
North Dakota *(2005–2007)*								
Across all grade levels					✱[2]	✱[2]	✱	→←
Elementary (grade 4)	→←[2, 3]	←→[2, 3]	→←[2, 3]	→←[2, 3]	←→	←→	←→	→←
Middle (grade 8)	←→[2]	←→[2]	←→[2]	→←[2]	→←	→←	→←	→←
High (grade 11)	←→[2, 3]	←→[2, 3]	→←[2]	→←[2]	→←[2]	→←[2]	→←	→←
Ohio *(2004–2007)*								
Across all grade levels	→←		→←[3]				→←[3]	
Elementary (grade 3)	→←	❍	→←	❍	←→[2]	❍	→←	❍
Middle (grade 8)	→←	❍	→←	❍	←→[2]	❍	→←	❍
High (OH Grad Test)	→←	❍	→←[3]	❍	←→[2, 3]	❍	→←[3]	❍
Oklahoma *(2002–2007)*								
Across all grade levels	✱		✱		✱		✱[1]	
Elementary (grade 5)	→←	❍	→←	❍	→←	❍	→←[1]	❍
Middle (grade 8)	→←	❍	→←	❍	→←	❍	→←[1]	❍
High (Algebra I)	←→	❍	←→	❍	←→	❍	←→[1]	❍
Oregon *(2002–2007)*								
Across all grade levels	→←[3]	✱[3]			✱[3]	✱[3]		✱[1]
Elementary (grade 5)	→←	←→	→←[3]	→←[3]	→←	→←	♦	→←[1]
Middle (grade 8)	→←	✪	→←[3]	✪[3]	→←[3]	✪[3]	♦	→←[1]
High (grade 10)	→←[3]	✪[3]	→←[3]	→←[3]	←→[3]	→←[3]	♦	←→[1]
Pennsylvania *(2002–2007)*								
Across all grade levels	→←[3]	→←[3]					→←[1, 3]	→←[1, 3]
Elementary (grade 5)	→←	→←	→←[3]	→←[3]	→←[2, 3]	→←[2, 3]	→←[1]	→←[1]
Middle (grade 8)	→←[3]	→←[3]	→←[3]	→←[3]	→←[2, 3]	→←[2, 3]	→←[1, 3]	→←[1, 3]
High (grade 11)	→←[3]	→←[3]	→←[3]	→←[3]	→←[2, 3]	→←[2, 3]	→←[1, 3]	→←[1, 3]
Rhode Island *(2004–2006)*								
Across all grade levels								
Elementary (grade 4)	♦	❍	♦	❍	♦	❍	♦	❍
Middle (grade 8)	♦	❍	♦	❍	♦	❍	♦	❍
High (grade 11)	→←	❍	→←	❍	→←	❍	❍	❍
South Carolina *(2002–2007)*								
Across all grade levels	✱	✱					✱[1, 3]	
Elementary (grade 4)	←→	✪	←→[3]	←→[3]	→←[2]	✪[2]	←→[1]	❍
Middle (grade 8)	←→	→←	←→[3]	←→[3]	✪[2]	→←[2]	←→[1]	❍
High (grade 10)	→←	→←	→←[3]	→←[3]	→←[2]	→←[2]	→←[1, 3]	→←[1, 3]

Continued on page 112

Table 5-P. State-by-State Summary of Achievement Gap Trends in Math (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
South Dakota *(2005–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	❍	♦	❍	♦	❍	♦	❍
Middle (grade 8)	♦	❍	♦	❍	♦	❍	♦	❍
High (grade 11)	♦	❍	♦	❍	♦	❍	♦	❍
Tennessee *(2004–2007)*								
Across all grade levels	→←	→←					→←[3]	→←[3]
Elementary (grade 4)	→←	→←	→←[3]	→←[3]	→←[2, 3]	→←[2, 3]	→←	→←
Middle (grade 8)	→←	→←	→←[3]	→←[3]	→←[2]	→←[2]	→←	→←
High (grade 9)	→←	→←	→←[3]	→←[3]	←→[2]	←→[2]	→←[3]	→←[3]
Texas *(2005–2007)*								
Across all grade levels	→←	✱	→←	→←	✱	✱	→←	→←
Elementary (grade 4)	→←	→←	→←	→←	←→	←→	→←	→←
Middle (grade 8)	→←	→←	→←	→←	→←	→←	→←	→←
High (grade 10)	→←	✪	→←	→←	←→	←→	→←	→←
Utah *(2004–2007)*								
Across all grade levels	✱[2]	←→[2]	✱	✱	→←[2, 3]	✱[2, 3]	→←	✱
Elementary (grade 4)	←→	←→	→←	→←	→←	✪	→←	→←
Middle (Pre-Algebra)	→←	←→	→←	←→	→←[3]	←→[3]	→←	→←
High (Geometry)	→←[2]	←→[2]	←→	←→	→←[2, 3]	←→[2, 3]	→←	←→
Vermont *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	❍	♦	❍	♦	❍	♦	❍
Middle (grade 8)	♦	❍	♦	❍	♦	❍	♦	❍
High (grade 10)	♦	❍	♦	❍	♦	❍	♦	❍
Virginia *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	❍	♦	❍	♦	❍	♦	❍
Middle (grade 8)	♦	❍	♦	❍	♦	❍	♦	❍
High (grade 11)	♦	❍	♦	❍	♦	❍	♦	❍
Washington *(2002–2007)*								
Across all grade levels	←→[3]	→←[3]			←→	→←		
Elementary (grade 4)	←→	→←	←→[3]	→←[3]	←→	→←	♦	♦
Middle (grade 7)	←→	→←	←→[3]	→←[3]	←→	→←	♦	♦
High (grade 10)	←→[3]	→←[3]	←→[3]	→←[3]	←→	→←	♦	♦

Continued on page 113

Table 5-P. State-by-State Summary of Achievement Gap Trends in Math (continued)

State, Years & Grade Levels Analyzed	African American/ White		Latino/White		Native American/ White		Low-Income/ Not Low-Income	
	PP	ES	PP	ES	PP	ES	PP	ES
West Virginia *(2004–2007)*								
Across all grade levels	→←	→←					→←	
Elementary (grade 4)	→←	→←	←→ [2, 3]	❍	→← [2]	❍	→←	❍
Middle (grade 8)	→←	→←	←→ [2, 3]	❍	←→ [2, 3]	❍	→←	❍
High (grade 10)	→←	→←	←→ [2, 3]	❍	←→ [2]	❍	→←	❍
Wisconsin *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 10)	♦	♦	♦	♦	♦	♦	♦	♦
Wyoming *(2006–2007)*								
Across all grade levels								
Elementary (grade 4)	♦	♦	♦	♦	♦	♦	♦	♦
Middle (grade 8)	♦	♦	♦	♦	♦	♦	♦	♦
High (grade 11)	♦	♦	♦	♦	♦	♦	♦	♦

Table reads: In Alabama, the gap between African American and white students narrowed at all three grade levels analyzed for this report according to the percentage of students scoring at the proficient level or above on the state math tests. There were not enough years (only 1-2 years) of effect size data to determine the gap trend according to this measure.

Note: Gap trends for students with disabilities and English language learners are not shown because state and federal policy changes may have affected the year-to-year comparability of test results for these subgroups. Trends for Asian students are not shown because in most states this subgroup performed as well as or better than white students.

[1] Low-income students were compared with all tested students in the state rather than with students who are not low-income.

[2] The number of students tested in this subgroup at this grade level was fewer than 500, so changes for this subgroup should be interpreted with caution.

[3] The number of students tested in this subgroup changed by at least +/-25% between the first and last year shown, so changes in results over time may be due to changes in the composition of the subgroup as well as changes in achievement.

CHAPTER 6

Comparing State Trends in Achievement Gaps with NAEP Gap Trends

Summary of Findings

In addition to comparing overall achievement trends on state tests and the National Assessment of Educational Progress, our study also compared trends in achievement gaps on state tests and NAEP. These comparisons focused on gaps in reading and math at the elementary and middle school levels for four subgroups—African American, Latino, and Native American students (all compared with white students) and students from low-income families (compared with students who are not from low-income families). Specifically, we compared changes in percentage proficient gaps on state tests between 2002 and 2007, or the closest span of years with comparable data, with changes in percentage basic gaps on NAEP between 2003 and 2007. These comparisons offer another perspective on whether the gap trends on state tests described in chapter 5 are corroborated by an independent measure of achievement. We arrived at several key findings:

- **Trends in achievement gaps on NAEP moved in the same direction as gap trends on state tests 60% of the time in states with sufficient data—a lower level of agreement than for overall achievement comparisons between NAEP and state tests.** We made 289 different comparisons of average yearly changes in percentage proficient gaps on state tests and percentage basic gaps on NAEP. We found that gaps narrowed on both NAEP and state tests about 53% of the time and widened on both about 8% of the time, for a total agreement rate of roughly 60%. This is less than the 81% agreement rate for the comparisons of overall achievement trends described in chapter 4. The other 40% of gap comparisons showed different trends on the two assessments.

- **Gaps trends on state tests and NAEP agreed much more often at the elementary level than at the middle school level, and more often for the Latino subgroup than for other subgroups.** Changes in gaps on state tests and NAEP moved in the same direction 73% of the time at the elementary/grade 4 level and 45% of the time at the middle school/grade 8 level, with variations by subgroup. Trends in gaps for the Latino subgroup moved in the same direction on state tests and NAEP 71% of the time. This is higher than the agreement rates of 62% for the low-income subgroup, 56% for African American students, and 48% for Native American students (but only a small number of states had sufficient data for Native Americans).

- **Achievement gaps narrowed on NAEP more often than they widened, but not as often as gaps narrowed on state tests.** In general, NAEP results paint a less positive picture of progress in narrowing gaps than state tests do. Of the 289 gap comparisons we made, gaps narrowed on NAEP 62% of the time while they narrowed on state tests 80% of the time. Still, a pattern of more gaps narrowing than widening emerged on both state tests and NAEP for most subgroups. An exception was in grade 8 math, where gaps on NAEP widened more often than they narrowed for most subgroups. Another exception was for Native American students, but too few states had sufficient data to discern a clear pattern for this subgroup.

- **In 2007, gaps between subgroups tended to be larger on NAEP than on state tests except in grade 8 math, but gaps remained sizeable on both assessments.** In reading at both grades 4 and 8, median achievement gaps were consistently larger on NAEP than on state tests. In grade 4 math, half of the gaps were also larger on NAEP. The major exception to the general pattern of larger NAEP gaps was in grade 8 math, where median gaps were consistently smaller on NAEP than on state tests. On both assessments, gaps between lower- and higher-performing subgroups often exceeded 20 percentage points in 2007.

- **Fewer states had sufficient data for gap comparisons with NAEP than for overall achievement comparisons with NAEP.** The number of states with sufficient data on both assessments to make gap comparisons ranged from a high of 30 states for the low-income subgroup in grade 4 reading to a low of 7 for the Native American subgroup in math. The numbers of states with sufficient data were smallest for the Native American subgroup (7 to 9 states, depending on the subject) and the Latino subgroup (10 to 16 states).

The next section of this chapter briefly explains the methods we used to compare gap trends on state tests and NAEP. The middle sections describe what we found when we analyzed the direction of gap trends (narrowed, widened, showed no change) on the two assessments and the size of gaps on both assessments in 2007. Tables are included for each type of comparison. A final section considers possible explanations for the similarities and differences between the gap trends on state tests and NAEP.

How We Compared Gap Trends on State Tests and NAEP

In addition to reporting overall achievement results for each state, the National Assessment of Educational Progress also reports state-by-state results for major subgroups of students. Using these NAEP subgroup data, we compared achievement gaps on NAEP with gap trends on state tests. For these gap comparisons, we adopted the same general approach and criteria that we used for the comparisons of overall achievement trends described in chapter 4.

For the state test side of the comparisons, we looked at changes in gaps between subgroups in terms of the percentages of students performing at or above the proficient level. The grade levels and span of years analyzed on state tests varied somewhat, depending on data availability, but were the same grades and years used for other analyses in this report—usually grades 4 and 8 (or an adjacent grade where sufficient comparable data were not available) and years 2002 through 2007 (or the closest span of years with comparable data).

For the NAEP side, we looked at changes in gaps between the 2003 and 2007 NAEP administrations in grades 4 and 8. After reviewing a range of evidence, we determined that it was more appropriate to compare percentages proficient on state tests with percentages of students performing at or above the basic level on NAEP, for reasons explained in more detail in box 4-A of chapter 4. The NAEP definition of "proficient," which is not tied to any specific state's curriculum standards, typically represents a more ambitious and aspirational concept of competent performance than the definitions of proficiency used by the vast majority of states (and by several other countries). State definitions of proficient are tied to the state's specific curriculum standards and are used for high-stakes accountability decisions under the No Child Left Behind Act, so these definitions may represent a more realistic vision of proficiency that takes into account students' current level of achievement.

We looked only at percentages proficient/basic because the number of comparisons involved in analyzing just this one indicator was already high. To add a second indicator of effect sizes would have made the amount of data overwhelming. In addition, fewer states had effect size data than had percentages proficient, and still fewer had effect size data broken down by subgroup. Thus, the numbers of states with sufficient data by subgroup would have been even smaller for effect sizes than for percentages proficient, limiting the conclusions we could draw.

Like other analyses in this report, our comparisons of gap trends focused on *average* yearly changes in gaps and included only states with at least three years of comparable state test data for a particular subgroup at a specific grade in reading or math. This approach ensured that we were looking at an amount of change reflecting a year's worth of progress on both state tests and NAEP, even when the two assessments covered different spans of years.

We did not analyze trends for subgroups that were small on state tests according to our definition (fewer than 500 students) or small on NAEP according to NAEP's reporting rule (fewer than 62 students).[4] Nor did we analyze trends for subgroups that had increased or decreased in size by 25% over the years analyzed. More details about the rules used to include subgroups can be found in chapter 2.

In essence, we were analyzing three layers of differences—the difference in test results between two subgroups of students (a gap), the difference in the size of this gap between a starting and ending year (a trend), and the difference in trends between state tests and NAEP. The outcomes of these types of multilayered comparisons are bound to be less precise than the outcome when just one type of comparison is involved.

Direction of Gap Trends

We sought to determine whether trends in achievement gaps on NAEP had moved in the same direction as or a different direction from gaps on state tests—in other words, whether gaps had narrowed, widened, or stayed the same on both types of assessments over the years analyzed. For each of the four subgroups mentioned above, we compared gap trends in four grade/subject combinations: elementary/grade 4 reading, middle school/grade 8 reading, elementary/grade 4 mathematics, and middle school/grade 8 math. We compared gap trends for the post-NCLB period analyzed on state tests with gap trends between 2003 and 2007 on NAEP.

NUMBER OF COMPARISONS AND AVAILABILITY OF DATA

Altogether, we did up to 16 different gap trend comparisons between state tests and NAEP in each state with sufficient data, for a total of 289 comparisons. The number of states with sufficient data varied from a high of 30 states for low-income students in grade 4 reading to a low of 7 states for Native American students in math. The number of states with sufficient data was smallest for the Native American subgroup (from 7 to 9 states, depending on the subject) and the Latino subgroup (from 10 to 16 states). These numbers were well short of 50 because many states had subgroups that were small according to our state test standard or the NAEP standard or that had changed significantly in size. In addition, some states lacked data for particular subgroups for other reasons. And several states lacked trend data for any of the analyses in this report because they had introduced new tests or changed their cut scores.

[4] NAEP's minimum subgroup size of 62 is much smaller than the minimum chosen for this study because NAEP exams are given to samples of students and results must be extrapolated to a whole state.

LOOKING AT GAP TRENDS ACROSS GRADES AND SUBJECTS

Table 6-A provides a broad picture of achievement gap trends on state tests and NAEP across all grades and subjects analyzed. The rows display the total number and percentage of gaps that narrowed, showed no change, or widened on each type of assessment over the periods analyzed. As table 6-A reveals, considerably more gaps narrowed on state tests than on NAEP—80% on state tests versus 62% on NAEP. Conversely, considerably more gaps widened on NAEP than on state tests—36% versus 14%, respectively.

Table 6-A. Summary of Gap Trends on State Tests and NAEP

	Number *(percentage)* of			
	→← Gaps that narrowed	✪ Gaps with no change	←→ Gaps that widened	Total
State tests	232 *(80%)*	16 *(6%)*	41 *(14%)*	289 *(100%)*
NAEP	178 *(62%)*	7 *(2%)*	104 *(36%)*	289 *(100%)*

Table reads: Of 289 achievement gap trends analyzed on state tests, 232 gaps (or 80% of the gaps) narrowed, 16 gaps (6%) showed no change, and 41 gaps (14%) widened.

Table 6-B shows how often trends moved in the same direction on state tests and NAEP for states with sufficient data on both assessments—in other words, how often gaps narrowed or widened on both. Essentially, this table summarizes key data from the more detailed tables for each subject, subgroup, and grade that follow in this section. Of the 289 gap comparisons made, gaps narrowed on both state tests and NAEP about 53% of the time and widened on both assessments about 8% of the time. Altogether, gap trends on the two assessments agreed about 60% of the time, when rounded.

When trends moved in different directions on state tests and NAEP, the NAEP trends were not always less positive. In some instances, gaps narrowed on NAEP but widened or showed no change on state tests.

The extent of agreement in gap trends between state tests and NAEP varied by grade level and subgroup, as displayed in **table 6-C**. Gap trends moved in the same direction on both assessments more often at the elementary/grade 4 level than at the middle school/grade 8 level—73% of the time for elementary and 45% of the time for middle school, with variations by subgroup. Among subgroups, the highest level of agreement between state gap trends and NAEP gap trends occurred for Latino students; trends for this subgroup agreed 71% of the time. The lowest rate of agreement, 48%, occurred for Native American students, but only a small number of states had sufficient data to count gap trends for this subgroup. Gap trends moved in the same direction 56% of the time for the African American subgroup and 62% of the time for the low-income subgroup.

Table 6-B. Agreement of Gap Trends Between State Tests and NAEP

Subject	NAEP Grade	Subgroup	Gap Narrowed on Both	Gap Widened on Both	Total Instances of Agreement	Number of States with Sufficient Data	Percentage of Gap Trends in Agreement
Reading	4	African American	15	1	16	26	62%
	4	Latino	10	1	11	16	69%
	4	Native American	4	0	4	8	50%
	4	Low-income	20	0	20	30	67%
	8	African American	9	1	10	25	40%
	8	Latino	6	0	6	11	55%
	8	Native American	1	3	4	9	44%
	8	Low-income	13	0	13	27	48%
Math	4	African American	19	2	21	26	81%
	4	Latino	15	0	15	15	100%
	4	Native American	2	2	4	7	57%
	4	Low-income	22	0	22	26	85%
	8	African American	5	4	9	23	39%
	8	Latino	4	1	5	10	50%
	8	Native American	1	2	3	7	43%
	8	Low-income	6	5	11	23	48%
Total			152	22	174	289	
Percentage of gaps in agreement			53%	8%			60%*

Table reads: Twenty-six states had sufficient data to analyze gaps on state tests and NAEP between African American and white students in elementary/grade 4 reading. In 15 of the 26 states, the African American-white gap in grade 4 reading narrowed on both state tests and NAEP during the period analyzed, while in 1 state this gap widened on both. On the whole, the African American-white gap in grade 4 reading moved in the same direction on state tests and NAEP in 62% of the instances analyzed.

Note: There were no instances of a gap showing no change on both state tests and NAEP.

* The figures in this row add up to more than 60% due to rounding.

Table 6-C. Agreement of Gap Trends Between State Tests and NAEP by Subject, Grade Level, and Subgroup

Type of Gap Trend	Number of Instances of Agreement	Total Number of Comparisons	Percentage of Gap Trends in Agreement
African American-white	56	100	56%
Latino-white	37	52	71%
Native American-white	15	31	48%
Low-income & non-low-income	66	106	62%
Elementary/grade 4	113	154	73%
Middle school/grade 8	61	135	45%
Reading	84	152	55%
Math	90	137	66%

Table reads: A total of 100 comparisons were made of trends in achievement gaps between African American and white students on state tests and NAEP. In 56 of these comparisons, trends on both assessments moved in the same direction (either both narrowing or both widening) for an agreement rate of 56%.

AFRICAN AMERICAN-WHITE GAP, GRADE 4 READING

As shown in **table 6-D**, 26 states had sufficient data to compare gaps between African American and white students in elementary/grade 4 reading. In 22 states, the African American-white gap narrowed on state tests, and in 15 of these states this gap also narrowed on NAEP. In 7 states, however, the black-white gap shrunk on state tests but grew on NAEP. In 1 state, this gap widened on both assessments. On the whole, there were more instances of widening black-white gaps on NAEP than on state tests, but both state tests and NAEP showed more instances of gaps narrowing than widening.

Table 6-D. Comparison of African American–White Gap Trends on State Tests and NAEP in Elementary/Grade 4 Reading

	Type of Trend	NAEP percentage basic			
		→← Gap narrowed	✪ No change in gap	←→ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 15 states: AL, AZ, FL, IA, IN, LA, MA, MD, NC, NJ, NV, OK, PA, TX, WV		→← ←→ 7 states: CO, KY, MS, NM, OR, TN, WA	22
	✪ No change in gap	✪ →← 2 states: CA, SC		✪ ←→ 1 state: AR	3
	←→ Gap widened			←→ ←→ 1 state: OH	1
	Total number of states	17	0	9	26

Table reads: Twenty-six states had sufficient data to compare gap trends for African American students on state elementary school reading tests and the NAEP grade 4 reading assessment. In 22 of these states, the reading achievement gap between African American and white students narrowed on state tests for the post-NCLB years with comparable data. Of the 22 states with narrowing trends on state tests, the African American-white gap also narrowed on NAEP in 15 states, while it widened on NAEP in 7 states.

Note: The following 24 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: AK, CT, DE, GA, HI, ID, IL, KS, ME, MI, MN, MO, MT, ND, NE, NH, NY, RI, SD, UT, VA, VT, WI, WY.

LATINO-WHITE GAP, GRADE 4 READING

In 10 of the 16 states with sufficient data for the Latino subgroup, the Latino-white gap narrowed in elementary/grade 4 reading on both state tests and NAEP, and in one state this gap widened on both assessments (see **table 6-E**). In 4 states with narrowing Latino-white gaps on state tests, this gap widened on NAEP. Another state with a narrowing gap on the state test showed no change on NAEP. Although the Latino-white gap grew smaller on both assessments in the majority of states with sufficient data, there were more instances of gaps widening on NAEP than on state tests.

Table 6-E. Comparison of Latino–White Gap Trends on State Tests and NAEP in Elementary/Grade 4 Reading

	Type of Trend	NAEP percentage basic: →← Gap narrowed	NAEP percentage basic: ✪ No change in gap	NAEP percentage basic: ←→ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 10 states: AZ, CA, FL, IA, ID, MA, NJ, OH, TX, UT	→← ✪ 1 state: NM	→← ←→ 4 states: CO, MD, NV, OK	15
	✪ No change in gap				0
	←→ Gap widened			←→ ←→ 1 state: AR	1
	Total number of states	10	1	5	16

Table reads: Sixteen states had sufficient data to compare gap trends for Latino students on state elementary school reading tests and the NAEP grade 4 reading assessment. In 15 of these states, the reading achievement gap between Latino and white students narrowed on state tests for the post-NCLB years with comparable data. Of the 15 states with narrowing trends on state tests, the Latino-white gap narrowed on NAEP in 10 states, showed no change on NAEP in 1 state, and widened on NAEP in 4 states.

Note: The following 34 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: AK, AL, CT, DE, GA, HI, IL, IN, KS, KY, LA, ME, MI, MN, MO, MS, MT, NC, ND, NE, NH, NY, OR, PA, RI, SC, SD, TN, VA, VT, WA, WI, WV, WY.

NATIVE AMERICAN-WHITE GAP, GRADE 4 READING

Only eight states had sufficient data to compare achievement gaps between Native American and white students in elementary/grade 4 reading. Usually this was because the number of Native American test-takers was too small to report reliable trends in most states. In four of the eight states with sufficient data, the Native American-white gap decreased on both state tests and NAEP (see **table 6-F**). In three states, the Native American-white gap moved in opposite directions, narrowing on state tests and widening on NAEP. In the remaining state, this gap widened on the state test but showed no change on NAEP.

Table 6-F. Comparison of Native American–White Gap Trends on State Tests and NAEP in Elementary/Grade 4 Reading

	Type of Trend	NAEP percentage basic: →← Gap narrowed	NAEP percentage basic: ✪ No change in gap	NAEP percentage basic: ←→ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 4 states: AZ, MT, NM, OK		→← ←→ 3 states: ND, SD, WA	7
	✪ No change in gap				0
	←→ Gap widened		←→ ✪ 1 state: AK		1
	Total number of states	4	1	3	8

Table reads: Eight states had sufficient data to compare gap trends for Native American students on state elementary school reading tests and the NAEP grade 4 reading assessment. In seven of these states, the reading achievement gap between Native American and white students narrowed on state tests for the post-NCLB years with comparable data. Of the seven states with narrowing trends on state tests, the Native American-white gap also narrowed on NAEP in four states, while it widened on NAEP in three states.

Note: The following 42 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: AL, AR, CA, CO, CT, DE, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MA, MD, ME, MI, MN, MO, MS, NC, NE, NH, NJ, NV, NY, OH, OR, PA, RI, SC, TN, TX, UT, VA, VT, WI, WV, WY.

LOW-INCOME AND NON-LOW-INCOME GAP, GRADE 4 READING

The gap in elementary/grade 4 reading between low-income students and students who are not low-income narrowed on both state tests and NAEP in the majority of the states with sufficient data for this subgroup. Of the 30 states with sufficient data, shown in **table 6-G**, 20 states reported that the gap for low-income students had narrowed on both assessments. In 8 states, trends on the two assessments diverged. In 6 of these 8 states, the low-income gap narrowed on state tests but widened on NAEP, while in the remaining 2 states, this gap narrowed on NAEP but widened on state tests.

Table 6-G. Comparison of Gap Trends for Low-Income and Non-Low-Income Students on State Tests and NAEP in Elementary/Grade 4 Reading

	Type of Trend	NAEP percentage basic			
		→← Gap narrowed	✪ No change in gap	←→ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 20 states: AK, AL, AZ, FL, GA, HI, IA, IN, LA, MD, MS, MT, NC, NJ, NM, NV, OH, OK, PA, UT	→← ✪ 1 state: KY	→← ←→ 6 states: AR, ID, SD, TN, TX, WV	27
	✪ No change in gap	✪ →← 1 state: CA			1
	←→ Gap widened	←→ →← 2 states: ND, SC			2
	Total number of states	23	1	6	30

Table reads: Thirty states had sufficient data to compare gap trends between low-income and non-low-income students on state elementary school reading tests and the NAEP grade 4 reading assessment. In 27 of these states, the reading achievement gap between low-income students and students who are not low-income narrowed on state tests for the post-NCLB years with comparable data. Of the 27 states with narrowing trends on state tests, the gap between low-income and non-low-income students narrowed on NAEP in 20 states, showed no change on NAEP in 1 state, and widened on NAEP in 6 states.

Note: The following 20 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: CO, CT, DE, IL, KS, MA, ME, MI, MN, MO, NE, NH, NY, OR, RI, VA, VT, WA, WI, WY.

AFRICAN AMERICAN-WHITE GAP, GRADE 8 READING

For the African American subgroup, state test score trends and NAEP trends in middle school/grade 8 reading often moved in different directions. Most state tests indicated that the African American-white gap had narrowed, while NAEP results showed an even split between narrowing and widening gaps. In 21 of the 25 states with sufficient data, the black-white gap narrowed on state middle school reading tests, but in only 9 states did this gap shrink on grade 8 NAEP as well (see **table 6-H**). By contrast, 11 of the states with narrowing gaps for African Americans on state tests had widening gaps on NAEP. And in 3 states, the black-white gap grew wider on state tests but shrunk on NAEP.

Table 6-H. Comparison of African American–White Gap Trends on State Tests and NAEP in Middle School/Grade 8 Reading

	Type of Trend	NAEP percentage basic: →← Gap narrowed	NAEP percentage basic: ✪ No change in gap	NAEP percentage basic: ↔ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 9 states: AR, FL, IA, KY, LA, NM, NV, OK, TX	→← ✪ 1 state: MD	→← ↔ 11 states: AL, IN, MS, NC, NJ, OH, OR, SC, TN, WA, WV	21
	✪ No change in gap				0
	↔ Gap widened	↔ →← 3 states: AZ, CO, MA		↔ ↔ 1 state: CA	4
	Total number of states	12	1	12	25

Table reads: Twenty-five states had sufficient data to compare gap trends for African American students on state middle school (usually grade 8) reading tests and the NAEP grade 8 reading assessment. In 21 of these states, the reading achievement gap between African American and white students narrowed on state tests for the post-NCLB years with comparable data. Of the 21 states with narrowing trends on state tests, the African American-white gap narrowed on NAEP in 9 states, showed no change on NAEP in 1 state, and widened on NAEP in 11 states.

Note: The following 25 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: AK, CT, DE, GA, HI, ID, IL, KS, ME, MI, MN, MO, MT, ND, NE, NH, NY, PA, RI, SD, UT, VA, VT, WI, WY.

LATINO-WHITE GAP, GRADE 8 READING

As **table 6-I** indicates, 11 states had sufficient data to examine gap trends for Latino students in middle school/grade 8 reading. In 6 of these 11 states, state tests and NAEP showed the same narrowing trend for the Latino-white gap. But in the remaining 5 states, Latino-white gap trends moved in different directions; this includes 3 states in which the gap narrowed on the state test but widened on NAEP.

Table 6-I. Comparison of Latino–White Gap Trends on State Tests and NAEP in Middle School/Grade 8 Reading

	Type of Trend	NAEP percentage basic			
		→← Gap narrowed	✪ No change in gap	←→ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 6 states: CO, IA, MA, NM, TX, UT		→← ←→ 3 states: ID, OH, OK	9
	✪ No change in gap			✪ ←→ 1 state: AZ	1
	←→ Gap widened	←→ →← 1 state: CA			1
	Total number of states	7	0	4	11

Table reads: Eleven states had sufficient data to compare gap trends for Latino students on state middle school (usually grade 8) reading tests and the NAEP grade 8 reading assessment. In nine of these states, the reading achievement gap between Latino and white students narrowed on state tests for the post-NCLB years with comparable data. Of the nine states with narrowing trends on state tests, the Latino-white gap narrowed on NAEP in six states, while it widened on NAEP in three states.

Note: The following 39 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: AK, AL, AR, CT, DE, FL, GA, HI, IL, IN, KS, KY, LA, MD, ME, MI, MN, MO, MS, MT, NC, ND, NE, NH, NJ, NV, NY, OR, PA, RI, SC, SD, TN, VA, VT, WA, WI, WV, WY.

NATIVE AMERICAN-WHITE GAP, GRADE 8 READING

Nine states had sufficient data to compare achievement gaps between Native American and white students in middle school/grade 8 reading. As displayed in **table 6-J**, trends for this subgroup showed relatively low agreement between the two assessments and several instances where gaps widened. However, the number of states with sufficient data was small. On state tests, trends in the Native American-white gap were split about evenly, with five states reporting that this gap had narrowed and four reporting that it had widened. But on NAEP, the Native American-white gap narrowed in just two states and widened in seven. In only one state did this gap shrink on both assessments, while in three states it grew on both.

Table 6-J. Comparison of Native American–White Gap Trends on State Tests and NAEP in Middle School/Grade 8 Reading

	Type of Trend	NAEP percentage basic			
		→← Gap narrowed	✪ No change in gap	←→ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 1 state: ND		→← ←→ 4 states: MT, NC, OK, WA	5
	✪ No change in gap				0
	←→ Gap widened	←→ →← 1 state: SD		←→ ←→ 3 states: AK, AZ, NM	4
	Total number of states	2	0	7	9

Table reads: Nine states had sufficient data to compare gap trends for Native American students on state middle school (usually grade 8) reading tests and the NAEP grade 8 reading assessment. In five of these states, the reading achievement gap between Native American and white students narrowed on state tests for the post-NCLB years with comparable data. Of the five states with narrowing trends on state tests, the Native American-white gap also narrowed on NAEP in one state, while it widened on NAEP in four states.

Note: The following 41 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: AL, AR, CA, CO, CT, DE, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MA, MD, ME, MI, MN, MO, MS, NE, NH, NJ, NV, NY, OH, OR, PA, RI, SC, TN, TX, UT, VA, VT, WI, WV, WY.

LOW-INCOME AND NON-LOW-INCOME GAP, GRADE 8 READING

In 14 of the 27 states with sufficient data in middle school/grade 8 reading, gap trends for the low-income subgroup differed on state tests and NAEP. Trends in low-income gaps moved in the same direction on both assessments in 13 states. In 8 states, however, the low-income gap shrunk on state tests but grew on NAEP (see **table 6-K**). The remaining 6 states had a mixture of no change on one assessment and a widening or narrowing trend on the other.

Table 6-K. Comparison of Gap Trends for Low-Income and Non-Low-Income Students on State Tests and NAEP in Middle School/Grade 8 Reading

	Type of Trend	NAEP percentage basic: →← Gap narrowed	NAEP percentage basic: ✪ No change in gap	NAEP percentage basic: ←→ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 13 states: AK, AL, FL, HI, IA, ID, MD, MT, NC, NJ, OH, OK, UT	→← ✪ 1 state: WV	→← ←→ 8 states: AR, KY, ND, NV, SC, SD, TN, TX	22
	✪ No change in gap	✪ →← 2 states: CA, GA		✪ ←→ 2 states: LA, NM	4
	←→ Gap widened		←→ ✪ 1 state: AZ		1
	Total number of states	15	2	10	27

Table reads: Twenty-seven states had sufficient data to compare gap trends for students from low-income families on state middle school (usually grade 8) reading tests and the NAEP grade 8 reading assessment. In 22 of these states, the reading achievement gap between low-income students and students who are not low-income narrowed on state tests for the post-NCLB years with comparable data. Of the 22 states with narrowing trends on state tests, the gap between low-income and non-low-income students narrowed on NAEP in 13 states, showed no change on NAEP in 1 state, and widened on NAEP in 8 states.

Note: The following 23 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: CO, CT, DE, IL, IN, KS, MA, ME, MI, MN, MO, MS, NE, NH, NY, OR, PA, RI, VA, VT, WA, WI, WY.

AFRICAN AMERICAN-WHITE GAP, GRADE 4 MATH

In elementary/grade 4 math, trends in the gap between African American and white students tended to move in the same direction on both state tests and NAEP. In 19 of the 26 states with sufficient data, the black-white gap narrowed on both state tests and NAEP, as shown in **table 6-L**. The black-white gap widened on both assessments in 2 states. The other 5 states showed divergent trends on state tests and NAEP.

Table 6-L. Comparison of African American–White Gap Trends on State Tests and NAEP in Elementary/Grade 4 Math

	Type of Trend	NAEP percentage basic: →← Gap narrowed	NAEP percentage basic: ✪ No change in gap	NAEP percentage basic: ↔ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 19 states: AL, AZ, CA, CO, FL, GA, IA, IN, LA, MA, MD, MS, NJ, NM, OH, OK, PA, TN, TX		→← ↔ 3 states: KY, OR, WV	22
	✪ No change in gap	✪ →← 1 state: NV			1
	↔ Gap widened	↔ →← 1 state: AR		↔ ↔ 2 states: SC, WA	3
	Total number of states	21	0	5	26

Table reads: Twenty-six states had sufficient data to compare gap trends for African American students on state elementary school math tests and the NAEP grade 4 math assessment. In 22 of these states, the math achievement gap between African American and white students narrowed on state tests for the post-NCLB years with comparable data. Of the 22 states with narrowing trends on state tests, the African American-white gap also narrowed on NAEP in 19 states, while it widened on NAEP in 3 states.

Note: The following 24 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: AK, CT, DE, HI, ID, IL, KS, ME, MI, MN, MO, MT, NC, ND, NE, NH, NY, RI, SD, UT, VA, VT, WI, WY.

LATINO-WHITE GAP, GRADE 4 MATH

Fifteen states had sufficient data to examine achievement gap trends between Latino and white students in elementary/grade 4 math. As **table 6-M** illustrates, the Latino-white gap narrowed on both state tests and NAEP in all 15 states—a perfect correspondence between state and NAEP results in a positive direction.

Table 6-M. Comparison of Latino–White Gap Trends on State Tests and NAEP in Elementary/Grade 4 Math

	Type of Trend	NAEP percentage basic			
		→← Gap narrowed	✪ No change in gap	←→ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 15 states: AZ, CA, CO, FL, IA, ID, MA, MD, NJ, NM, NV, OH, OK, TX, UT			15
	✪ No change in gap				0
	←→ Gap widened				0
	Total number of states	15	0	0	15

Table reads: Fifteen states had sufficient data to compare gap trends for Latino students on state elementary school math tests and the NAEP grade 4 math assessment. In all of these states, the math achievement gap between Latino and white students narrowed on state tests for the post-NCLB years with comparable data and also narrowed on NAEP. None of these 15 states showed no change in or a widening of the Latino-white gap on either measure.

Note: The following 35 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: AK, AL, AR, CT, DE, GA, HI, IL, IN, KS, KY, LA, ME, MI, MN, MO, MS, MT, NC, ND, NE, NH, NY, OR, PA, RI, SC, SD, TN, VA, VT, WA, WI, WV, WY.

NATIVE AMERICAN-WHITE GAP, GRADE 4 MATH

Just seven states had sufficient data to compare gap trends for Native American and white students in elementary/grade 4 math. In two of these states, the Native American-white gap narrowed on both state tests and NAEP, and in two states it widened on both assessments (see **table 6-N**). In the remaining three states, this gap narrowed on NAEP but showed no change or a widening on state tests. However, too few states had sufficient data to discern an overall pattern for the Native American subgroup.

Table 6-N. Comparison of Native American–White Gap Trends on State Tests and NAEP in Elementary/Grade 4 Math

	Type of Trend	NAEP percentage basic			
		→← Gap narrowed	✪ No change in gap	←→ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 2 states: AZ, OK			2
	✪ No change in gap	✪ →← 2 states: MT, NM			2
	←→ Gap widened	←→ →← 1 state: ND		←→ ←→ 2 states: AK, WA	3
	Total number of states	5	0	2	7

Table reads: Seven states had sufficient data to compare gap trends for Native American students on state elementary school math tests and the NAEP grade 4 math assessment. In two of these states, the math achievement gap between Native American and white students narrowed on state tests for the post-NCLB years with comparable data and also narrowed on NAEP. None of the states with narrowing trends on state tests showed no change or widening gaps on NAEP.

Note: The following 43 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: AL, AR, CA, CO, CT, DE, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MA, MD, ME, MI, MN, MO, MS, NC, NE, NH, NJ, NV, NY, OH, OR, PA, RI, SC, SD, TN, TX, UT, VA, VT, WI, WV, WY.

LOW-INCOME AND NON-LOW INCOME GAP, GRADE 4 MATH

In general gap trends for the low-income subgroup followed the same positive direction on both assessments in elementary/grade 4 math. In 22 of the 26 states with sufficient gap data, the gaps between low-income and non-low-income students narrowed on both state tests and NAEP. As illustrated in **table 6-O**, very few states reported that the low-income gap had widened—just 3 states on the state test and 1 state on NAEP. No state reported a flat trend for this gap.

Table 6-O. Comparison of Gap Trends for Low-Income and Non-Low-Income Students on State Tests and NAEP in Elementary/Grade 4 Math

		NAEP percentage basic			
	Type of Trend	→← Gap narrowed	✪ No change in gap	←→ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 22 states: AK, AL, AR, AZ, CA, CO, FL, HI, IA, ID, KY, LA, MD, NJ, NM, NV, OH, OK, PA, TN, TX, UT		→← ←→ 1 state: WV	23
	✪ No change in gap				0
	←→ Gap widened	←→ →← 3 states: MT, ND, SC			3
	Total number of states	25	0	1	26

Table reads: Twenty-six states had sufficient data to compare gap trends for students from low-income families on state elementary school math tests and the NAEP grade 4 math assessment. In 23 of these states, the math achievement gap between low-income students and students who are not low-income narrowed on state tests for the post-NCLB years with comparable data. Of the 23 states with narrowing trends on state tests, the gap between low-income and non-low-income students also narrowed on NAEP in 22 states, while it widened on NAEP in 1 state.

Note: The following 24 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: CT, DE, GA, IL, IN, KS, MA, ME, MI, MN, MO, MS, NC, NE, NH, NY, OR, RI, SD, VA, VT, WA, WI, WY.

AFRICAN AMERICAN-WHITE GAP, GRADE 8 MATH

In the 23 states with sufficient data, gap trends for African American students in middle school/grade 8 math were often at variance between state tests and NAEP (see **table 6-P**). Although the African American-white gap shrunk on state tests in 16 states, it also shrank on NAEP in just 5 of these states. Furthermore, this gap widened on NAEP more often than it narrowed—an opposite pattern from the state test data.

Table 6-P. Comparison of African American–White Gap Trends on State Tests and NAEP in Middle School/Grade 8 Math

	Type of Trend	NAEP percentage basic: →← Gap narrowed	NAEP percentage basic: ✪ No change in gap	NAEP percentage basic: ↔ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 5 states: CO, FL, LA, OK, OR	→← ✪ 1 state: AZ	→← ↔ 10 states: AL, IA, IN, MS, NJ, NV, OH, TN, TX, WV	16
	✪ No change in gap				0
	↔ Gap widened	↔ →← 3 states: KY, NM, SC		↔ ↔ 4 states: AR, MA, MD, WA	7
	Total number of states	8	1	14	23

Table reads: Twenty-three states had sufficient data to compare gap trends for African American students on state middle school (usually grade 8) math tests and the NAEP grade 8 math assessment. In 16 of these states, the math achievement gap between African American and white students narrowed on state tests for the post-NCLB years with comparable data. Of the 16 states with narrowing trends on state tests, the African American-white gap narrowed on NAEP in 5 states, showed no change on NAEP in 1 state, and widened on NAEP in 10 states.

Note: The following 27 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: AK, CA, CT, DE, GA, HI, ID, IL, KS, ME, MI, MN, MO, MT, NC, ND, NE, NH, NY, PA, RI, SD, UT, VA, VT, WI, WY.

LATINO-WHITE GAP, GRADE 8 MATH

Ten states had sufficient data in middle school/grade 8 math to compare Latino-white gap trends on both state tests and NAEP. As displayed in **table 6-Q**, the results corresponded on the two assessments in just half of these states, and the state tests gave a more positive picture than NAEP. In the eight states with narrowing Latino-white gaps on state tests, only four also showed narrowing gaps on NAEP, while the remaining four had widening gaps on NAEP. Gaps grew wider on both assessments in one state. And in one state, the Latino-white gap grew wider on the state test but narrower on NAEP.

Table 6-Q. Comparison of Latino–White Gap Trends on State Tests and NAEP in Middle School/Grade 8 Math

	Type of Trend	NAEP percentage basic: →← Gap narrowed	NAEP percentage basic: ✪ No change in gap	NAEP percentage basic: ←→ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 4 states: IA, ID, OH, UT		→← ←→ 4 states: AZ, CO, OK, TX	8
	✪ No change in gap				0
	←→ Gap widened	←→ →← 1 state: NM		←→ ←→ 1 state: MA	2
	Total number of states	5	0	5	10

Table reads: Ten states had sufficient data to compare gap trends for Latino students on state middle school (usually grade 8) math tests and the NAEP grade 8 math assessment. In eight of these states, the math achievement gap between Latino and white students narrowed on state tests for the post-NCLB years with comparable data. Of the eight states with narrowing trends on state tests, the Latino-white gap also narrowed on NAEP in four states, while it widened on NAEP in four states.

Note: The following 40 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: AK, AL, AR, CA, CT, DE, FL, GA, HI, IL, IN, KS, KY, LA, MD, ME, MI, MN, MO, MS, MT, NC, ND, NE, NH, NJ, NV, NY, OR, PA, RI, SC, SD, TN, VA, VT, WA, WI, WV, WY.

NATIVE AMERICAN-WHITE GAP, GRADE 8 MATH

Only seven states had sufficient data to compare gaps trends between Native American and white students in middle school/grade 8 math. In just one state, the Native American-white gap narrowed on both the state test and NAEP, and in two states, it widened on both assessments (see **table 6-R**). In the other four states, gap trends for Native Americans differed on the two assessments.

Table 6-R. Comparison of Native American–White Gap Trends on State Tests and NAEP in Middle School/Grade 8 Math

	Type of Trend	NAEP percentage basic: →← Gap narrowed	NAEP percentage basic: ✪ No change in gap	NAEP percentage basic: ↔ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 1 state: OK		→← ↔ 2 states: MT, ND	3
	✪ No change in gap			✪ ↔ 1 state: AZ	1
	↔ Gap widened	↔ →← 1 state: NM		↔ ↔ 2 states: AK, WA	3
	Total number of states	2	0	5	7

Table reads: Seven states had sufficient data to compare gap trends for Native American students on state middle school (usually grade 8) math tests and the NAEP grade 8 math assessment. In three of these states, the math achievement gap between Native American and white students narrowed on state tests for the post-NCLB years with comparable data. Of the three states with narrowing trends on state tests, the Native American-white gap also narrowed on NAEP in one state, while it widened on NAEP in two states.

Note: The following 43 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: AL, AR, CA, CO, CT, DE, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MA, MD, ME, MI, MN, MO, MS, NC, NE, NH, NJ, NV, NY, OH, OR, PA, RI, SC, SD, TN, TX, UT, VA, VT, WI, WV, WY.

LOW-INCOME AND NON-LOW-INCOME GAP, GRADE 8 MATH

Gap trends for low-income students in middle school/grade 8 math showed markedly different patterns on state tests and NAEP in the 23 states with sufficient data. As **table 6-S** illustrates, the gap between low-income and non-low-income students narrowed on state tests in 15 states, but also narrowed on NAEP in just 6 of these states. In 16 states, the low-income gap widened on NAEP, including 5 states in which this gap widened on both assessments. Indeed, the low-income gap at this grade widened on NAEP more than twice as often as it narrowed.

Table 6-S. Comparison of Gap Trends for Low-Income and Non-Low-Income Students on State Tests and NAEP in Middle School/Grade 8 Math

	Type of Trend	NAEP percentage basic			
		→← Gap narrowed	✪ No change in gap	←→ Gap widened	Total number of states
State test percentage proficient	→← Gap narrowed	→← →← 6 states: FL, HI, IA, ND, OK, WV		→← ←→ 9 states: AK, AL, ID, NJ, NV, OH, TN, TX, UT	15
	✪ No change in gap	✪ →← 1 state: LA		✪ ←→ 2 states: AZ, MD	3
	←→ Gap widened			←→ ←→ 5 states: AR, KY, MT, NM, SC	5
	Total number of states	7	0	16	23

Table reads: Twenty-three states had sufficient data to compare gap trends for students from low-income families on state middle school (usually grade 8) math tests and the NAEP grade 8 math assessment. In 15 of these states, the math achievement gap between low-income students and students who are not low-income narrowed on state tests for the post-NCLB years with comparable data. Of the 15 states with narrowing trends on state tests, the gap between low-income and non-low-income students also narrowed on NAEP in 6 states, while it widened on NAEP in 9 states.

Note: The following 27 states were not included in this comparison with NAEP because data from state tests or NAEP were unavailable or covered too few years to constitute a trend: CA, CO, CT, DE, GA, IL, IN, KS, MA, ME, MI, MN, MO, MS, NC, NE, NH, NY, OR, PA, RI, SD, VA, VT, WA, WI, WY.

Size of Achievement Gaps on NAEP and State Tests

In addition to comparing the direction of trends in achievement gaps on state tests and NAEP, we also looked at the size of the gaps between subgroups of students on the two assessments in 2007 (or 2006 for the three states that changed their tests significantly in 2007). In particular, we examined the size of the median gaps across all states on state tests and NAEP, as well as the smallest and largest gaps on the two assessments in any state.

Table 6-T gives a snapshot of the size of the gaps for each of the four subgroups analyzed in reading and math at elementary school/grade 4 and middle school/grade 8. In each of the grade and subject combinations, the table shows the median gap for a particular subgroup across all states with sufficient data. To give a sense of the range, the table also includes the smallest gap for that subgroup in any state (the minimum) and the largest gap for that subgroup in any state (the maximum). All gaps are expressed as percentage point differences—in other words, the difference between subgroups in the percentages of students performing at the proficient level on state tests or in the percentages basic on NAEP.

States were included in this analysis of gap size only if they had sufficient data to calculate gaps on both their state test and NAEP. Subgroups were not included in a given state if they were small according to our definition or NAEP's reporting rule.

The median gaps between subgroups were large on both state tests and NAEP—often exceeding 20 percentage points. For example, on state tests the median African American-white gap was 23 percentage points in grade 4 reading, 26 points in grade 8 reading, 24 points in grade 4 math, and 32 points in grade 8 math. On NAEP, the median African-American white gap was 30 percentage points in grade 4 reading, 28 points in grade 8 reading, 26 points in grade 4 math, and 29 points in grade 8 math.

The median gaps were consistently larger on NAEP than on state tests in reading, most notably in grade 4 reading. For example, in grade 4 reading the median gap was higher on NAEP than on state tests by 7 percentage points for the African American-white gap, 6 points for the Latino-white gap, 5 points for the Native American-white gap, and 10 points for the low-income and non-low-income gap. In grade 8 reading, the median gaps on NAEP were one or two percentage points higher than the median gaps on state tests. In grade 4 math, the pattern was mixed. NAEP gaps were larger for the African American-white and Latino-white gap comparisons, but state gaps were larger for the Native American-white gap. For the low-income and non-low-income gap, the gaps were the same on NAEP and state tests.

The major exception to the general pattern of larger gaps on NAEP was in grade 8 math, where the median gaps were consistently smaller on NAEP than on state tests.

The minimum and maximum gaps also reveal some interesting findings. No state had totally closed the gap for any of the four subgroups; the smallest minimum gap for any subgroup was 5 percentage points. On the other end of the scale, the maximum gaps in some states were quite large—as high as 54 percentage points between African American and white students on one state's grade 8 reading test. In several cases, the maximum gaps for various subgroups amounted to 30 or 40 percentage points or more. These great disparities in performance among subgroups indicate that several states have a long way to go to close achievement gaps.

Table 6-T. Magnitude of Achievement Gaps on State Tests and NAEP, 2007*
(Percentage Point Differences Between Two Subgroups)

Legend

State tests = Percentage point gaps in percentages proficient for 2007*

NAEP = Percentage point gaps in percentages basic for 2007

	Grade 4 Reading							
	African American–White		Latino–White		Native American–White		Low-Income–Not Low-Income	
	State	NAEP	State	NAEP	State	NAEP	State	NAEP
Median	23	30	20	26	25	30	18	28
Minimum	6	18	11	8	8	12	8	18
Maximum	36	42	40	38	33	44	38	40
	Grade 8 Reading							
	African American–White		Latino–White		Native American–White		Low-Income–Not Low-Income	
	State	NAEP	State	NAEP	State	NAEP	State	NAEP
Median	26	28	23	24	25	27	21	22
Minimum	7	17	10	13	7	13	9	14
Maximum	54	42	37	39	38	39	34	32
	Grade 4 Math							
	African American–White		Latino–White		Native American–White		Low-Income–Not Low-Income	
	State	NAEP	State	NAEP	State	NAEP	State	NAEP
Median	24	26	17	19	27	25	18	18
Minimum	9	18	6	9	8	8	5	11
Maximum	38	45	30	31	33	35	32	29
	Grade 8 Math							
	African American–White		Latino–White		Native American–White		Low-Income–Not Low-Income	
	State	NAEP	State	NAEP	State	NAEP	State	NAEP
Median	32	29	25	24	29	27	25	24
Minimum	13	11	10	16	8	8	8	14
Maximum	44	45	35	40	41	32	34	35

Table reads: On state reading tests for the elementary grade analyzed (usually grade 4), the median gap in percentages proficient between African American and white students was 23 percentage points in 2007. On the NAEP 2007 reading assessment for grade 4, the median gap in percentages basic between African American and white students was 30 percentage points.

*In the three states (HI, KY, OR) that made major changes in their testing programs in 2007, the state test data were from 2006 instead of 2007.

Possible Explanations for Similarities and Differences Between State Tests and NAEP

Both NAEP and state tests showed more gaps narrowing than widening. Some possible reasons for this improvement are summarized near the end of chapter 5. These include the possibilities that students in targeted subgroups may be receiving better instruction and learning more than in the past, and that student subgroups are benefiting from the interventions concentrated in urban school districts under the No Child Left Behind Act. Subtle manipulation of cut scores could be a factor in gaps narrowing on state tests, but that does not explain why such narrowing would also show up on NAEP.

We found that in 60% of the cases we analyzed, state and NAEP gap trends moved in the same direction within a state, bolstering the validity of the state test findings. If gaps narrowed on both state tests and NAEP, this provides more evidence of a real increase in learning among students in targeted subgroups. Likewise, if gaps widened on both assessments, this suggests a reason for concern that subgroups are not receiving the quality of instruction and resources they need.

In 40% of cases, state and NAEP gap trends diverged—most often because gaps narrowed on state tests but showed no change or widened on NAEP. Several factors may explain why state gap results are more positive, and some of these overlap with the possible reasons offered at the end of chapter 4 for inconsistent trends in overall achievement on state tests and NAEP:

- **Preparation for state tests**. Educators are adjusting their instruction, in both appropriate and inappropriate ways, to help students succeed on state tests. These practices can increase students' familiarity with their particular state's test and can lead to score inflation. Educators have less reason to engage in these practices to prepare students for NAEP. Only a sample of students and schools even takes NAEP, and no scores are reported at the school or student level.

- **Instructional sensitivity.** Under No Child Left Behind, state tests must be aligned with state standards outlining the academic content students should learn at a particular grade. If students learn the content embodied in their state's standards, these gains in learning are likely to show up as improvements in state test scores to at least some degree. NAEP, by contrast, is not deliberately aligned to any particular state's standards and may not assess what students are actually being taught. Gains in learning may not register as improvements on NAEP if the content of NAEP does not match what students were taught. For this reason, state tests are likely to be a better measure than NAEP of what is taught in most classrooms and may reflect larger gains. Because it is difficult for students to register improvement on NAEP, tiny gains may end up widening gaps almost as often as narrowing gaps, depending on which subgroup has shown improvement.

- **Motivation.** Both teachers and students are likely to be more motivated to ensure students perform well on state tests. Not only are state test scores publicly reported, but they are also used to determine which schools are targeted for sanctions under No Child Left Behind or the state accountability system. If low performance in a school persists, principals and teachers could be replaced. In some states, the same tests used for NCLB accountability also determine whether students will graduate or be promoted to the next grade. In these situations, students have a strong motivation to do well, and educators and policymakers have reason to focus remediation and prevention efforts on lower-performing subgroups to ensure students do not fail the exams. No similar incentives exist for NAEP because no consequences for students, teachers, or administrators are attached to NAEP scores.

- **Efforts to help subgroups meet AYP targets.** For a school or district to make adequate yearly progress under NCLB, all major subgroups must meet state test score targets. Schools have instituted various interventions and intensive instruction specifically to raise scores on state tests for subgroups that have not made AYP. These efforts could help to close gaps on state tests but may have limited impact on NAEP gaps.

- **Different grades and years analyzed.** We sometimes analyzed different grades and years on state tests and NAEP, which could produce somewhat different trends.

- **Different performance of subgroups relative to cut score.** For reasons explained in appendix 2, the size of a gap between two subgroups can depend on where the cut score is set for proficient performance (or basic performance in the case of NAEP). The gap between African American and white students, for example, tends to be larger if the cut score is set near the middle of the scale, closer to the mean, than at either extreme. Small changes over time in the distribution of scores for either or both subgroups involved in a gap comparison could have a different impact on the size of the gap depending on where cut scores are set relative to the mean on state tests and NAEP.

References

Achieve. (n.d.). *2005 NAEP results: State vs. nation*. Retrieved February 26, 2007, from www.achieve.org/node/482

Anderson M.A., Kaufman J, Simon T.R., Barrios L., Paulozzi L., Ryan G., et al. (2001). School-associated violent deaths in the United States, 1994–1999. *Journal of the American Medical Association, 286*(21), 2695–2702. Retrieved April 28, 2008, from www.cdc.gov/ncipc/schoolviolencejoc11149.pdf

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148. Retrieved April 22, 2008, from www.pdkintl.org/kappan/kbla9810.htm

Brophy, J., & Ames, C. (2005). NAEP testing for twelfth graders: Motivational issues. A paper prepared for the National Assessment Governing Board, August.

Center on Education Policy. (2005). *From the capital to the classroom: Year 3 of the No Child Left Behind Act*. Washington, DC: Author.

Center on Education Policy. (2006). *From the capital to the classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Author.

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

Center on Education Policy. (2007b). *Choices, changes, and challenges: Curriculum and instruction in the NCLB era*. Washington, DC: Author.

Center on Education Policy. (2008). *Instructional time in elementary schools: A closer look at changes for specific subjects*. Washington, DC: Author.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Cronin, J., Dahlin, M., Adkins, D., & Kingsbury, G.G. (2007). *The proficiency illusion*. Washington DC: Thomas B. Fordham Institute.

Dillon, S. (2005, November 26). Students ace state tests, but earn D's from U.S. *New York Times*. Retrieved May 2, 2007, from www.nytimes.com/2005/11/26/education/26tests.html?ei=5088&en=fdf05ea7edbf1440&ex=1290661200&partner=rssnyt&emc=rss&pagewanted=print

The Education Trust (2005). *Stalled in secondary: A look at student achievement since the No Child Left Behind Act*. Washington DC: author.

Fuller, B., Gesicki, K., Kang, E., & Wright J. (2006). *Is the No Child Left Behind Act working? The reliability of how states track achievement*. Berkeley, CA: Policy Analysis for California Education. Retrieved May 12, 2008, from http://pace.berkeley.edu/reports/WP.06-1.pdf

Hall, D., & Kennedy, S. (2006). *Primary progress, secondary challenge: A state by state look at student achievement patterns*. Washington, DC: The Education Trust. Retrieved May 12, 2008, from www2.edtrust.org/NR/rdonlyres/15B22876-20C8-47B8-9AF4-FAB148A225AC/0/PPSCreport.pdf

Hamilton, L.S., Stecher, B.M., Marsh, J.A., McCombs, J.S., Robyn, A., Russell, J.L., et al. (2007). *Standards-based accountability under No Child Left Behind*. Santa Monica, CA: RAND Corp.

Holland, P. (2002). Two indicators of change in the gaps between the CDFs of test score distributions. *Journal of Educational and Behavioral Statistics, 27*(1), 3-17.

Koretz, D. (2005.) Alignment, high stakes, and the inflation of test scores. In J. Herman & E. Haertel (Eds.), *Uses and misuses of data in accountability testing: Yearbook of the National Society for the Study of Education, 104* (2), 99-188. Malden, MA: Blackwell Publishing.

Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcome trends*. Cambridge, MA: Harvard Civil Rights Project.

Lepper, M. R., & Hederlong, J. (2000). Turning "play" into "work" and "work" into "play": 25 years of research on intrinsic versus extrinsic motivation. In C. Sansone & J. M. Harackewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance*. San Diego, CA: Academic Press.

Levitt, S.D., & Fryer, R.G. (2004). Falling behind: New evidence on the black-white achievement gap. *Education Next 4*(4). Retrieved April 22, 2008, from www.hoover.org/publications/ednext/3259506.html

Linn, R. (2000). Assessment and accountability. *Educational Researcher, 29*(2), 4-16.

National Center for Education Statistics. (2004). *Highlights from the Trends in International Mathematics and Science Study: TIMSS 2003.* Retrieved April 4, 2008, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2005005

National Center for Education Statistics. (2005). Common Core of Data. Retrieved on May 14, 2008, from http://nces.ed.gov/ccd/bat/

National Center for Education Statistics. (2006). *The nation's report card: Science 2005.* Retrieved June 6, 2006, from http://nces.ed.gov/nationsreportcard/pubs/main2005/2006466.asp

National Center for Education Statistics. (2007a). *Highlights from PISA 2006: Performance of U.S. 15-year-old students in science and mathematics literacy in an international context*. Retrieved April 4, 2008, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2008016

National Center for Education Statistics. (2007b). *Mapping 2005 state proficiency standards onto the NAEP scales*. NCES 2007-482. Washington, DC: U.S. Department of Education.

National Center for Education Statistics. (2007c). *The nation's report card: 12th-grade reading and mathematics 2005.* Retrieved June 6, 2006, from http://nces.ed.gov/nationsreportcard/pubs/main2005/2007468.asp

National Center for Education Statistics. (n.d.) *The NAEP reading achievement levels and the NAEP mathematics achievement levels*. Retrieved May 1, 2008, from http://nces.ed.gov/nationsreportcard/reading/achieve.asp

Peterson, P.E., & Hess, F.M. (2006). Keeping an eye on state standards: A race to the bottom? *Education Next*, 2006(3). Retrieved May 12, 2008 from http://media.hoover.org/documents/ednext20063_28.pdf.

Peterson, P. E., & West, M.R. (2003). The politics and practice of accountability. In Peterson, P. E., & West, M. R. (Eds.), *No Child Left Behind: The politics and practice of accountability*, Washington, DC: Brookings.

Phillips, G.W. (2007). *Linking NAEP achievement levels to TIMSS*. Washington, DC: American Institutes for Research.

Snipes, J., Horwitz, A., Soga, K., & Casserly, M. (2008). *Beating the odds, an analysis of student performance and achievement gaps on state assessments: Results from the 2006-07 school year*. Washington DC: Council of the Great City Schools. Retrieved May 9, 2008, from www.cgcs.org/pdfs/BTO8_Analysis.pdf

Stecher, B.M., & Hamilton, L.S. (2002, Spring). Putting theory to the test: Systems of "educational accountability" should be held accountable. *Rand Review, 26*(1), 17-23. Retrieved May 14, 2008, from http://rand.org/publications/randreview/issues/rr.04.02/theory.html

Stockburger, D. W. (1998). *Introductory statistics: Concepts, models, and applications*. Retrieved March 9, 2007, from www.psychstat.missouristate.edu/introbook/SBK11.htm

Stulich, S., Eisner, E., & McCrary, J. (2007). *National assessment of Title I*. Washington, DC: U.S. Department of Education.

Surgeon General of the United States. (2001). *Youth violence: A report of the Surgeon General*. Washington DC: U.S. Department of Health and Human Services. Retrieved April 28, 2008, from www.surgeongeneral.gov/library/youthviolence/default.htm

U.S. Census Bureau. (n.d.). U.S. Census 2000. Retrieved June 6, 2008, from www.census.gov/main/www/cen2000.html

Appendix 1. Study Methods

The Center on Education Policy designed the methods for conducting this study with advice from an expert panel and with close involvement and extensive technical support from our contractor, HumRRO. Many aspects of the study methods are described in the body of this report, particularly in chapter 2. Additional information about methods is provided here, as well as a list of study features that are new or improved this year.

New or Improved Features of This Year's Study

This year's study continues, updates, and refines the work on achievement that CEP initiated last year. It includes the following new features or improvements:

- Adding test data from school year 2006-07 and updating trends from 2002 through 2007, where comparable data are available
- Giving effect size equal prominence as an indicator of achievement along with the indicator of percentages of students scoring at or above the proficient level
- Doing additional comparisons of trends in state test scores with trends on the National Assessment of Educational Progress (NAEP), including new comparisons of the direction and magnitude of achievement gaps on NAEP and state tests
- Conducting new analyses of achievement gaps on state tests in one grade at each of three levels (elementary, middle, and high school), as well as trends across grade levels
- Posting all raw data collected from states for all grades on CEP's Web site, in addition to the state profiles posted online this year and last year

Collecting and Verifying Data

Data for the study were collected by HumRRO from July 2007 through April 2008. All 50 states provided at least some of the information we sought. As explained below, officials from each state were asked to verify the accuracy of the state information.

TYPES OF DATA COLLECTED

For the report on achievement that CEP published last year, HumRRO sought test-related information from every state for each year between 1999 and 2006. For this second annual report, HumRRO updated the data files with missing information from those years and added information from 2007, including the following:

1. Name of test(s) used for accountability
2. Testing contractor/developer with point-of-contact information
3. Independent quality assurance/verification contractor (if any) with point-of-contact information
4. Key state department of education staff (1-2 people, such as the testing director) with contact information
5. Test type (norm-referenced, criterion-referenced, augmented criterion-reference, computerized adaptive testing, other)
6. Test scoring scale
7. Description of and cut scores for performance levels on the test
8. Longitudinal/vertical scale or separate scale by test level
9. Grades tested in reading and mathematics
10. Grades included in determinations of percentages proficient in reading and math for adequate yearly progress under NCLB
11. Subjects tested other than reading and math (just a list with tested grade levels)
12. Timing of test (spring or fall)
13. Frequency of testing (usually annual)
14. Item types (only for reading and math tests used to calculate AYP)
15. Equating methodologies (summarized by type, such as Stocking-Lord, Rasch) and length of time this method has been used
16. Most current state-level mean score, number of test-takers (N-count), and standard deviations (SDs) for reading and math at all tested grades
17. Mean scores, N-counts, and SDs for all subgroups reported by the state
18. Mean scores and SDs for the entire state and by subgroup for each comparable year back to 1999
19. Percentages of students scoring in each proficiency category reported by the state for the entire state and by subgroup, at all grade levels for each comparable year back to 1999
20. Start date (if longitudinal scale is present)
21. Major changes to the tests from 2002 to the present—for example, changes in test, standards setting, cut scores, or tested grade levels
22. Studies of alignment of test to state content standards
23. Total K-12 public school enrollment in the 2006-07 school year

INITIAL DATA COLLECTION

A senior HumRRO project member revised templates and instructions for the data collection. She collected and organized information (such as specific Web site addresses) to streamline the data collection process. Using these templates and instructions, she trained staff members who were assigned to collect the data.

As a first step in collecting the desired data, HumRRO staff searched Web sites maintained by state departments of education. When Web sites lacked crucial information, HumRRO staff attempted to directly contact state department of education staff by e-mail or phone.

The data from each state were put into one descriptive MS Word table and two MS Excel file with one worksheet per assessment year. One Excel file contained results by achievement level and the other contained mean scale scores, standard deviations, and numbers of test-takers.

ANALYSIS OF PRELIMINARY DATA

HumRRO staff analyzed overall achievement trends using two indicators (percentage proficient and effect size computed from mean scale scores with standard deviations) for all 50 states. A two-page summary was prepared for state review that indicated the years believed to be comparable as well as the overall trends in achievement. Proficiency data were provided in graph form while mean scale scores and standard deviations were presented in tables. Introductory text described the rationale for the data interpretation and explained that the trends presented would be reported unless state personnel provided contradictory guidance.

STATE VERIFICATION PROCESS

HumRRO staff reviewed logs maintained during last year's data collection cycle to identify sources of confusion, misunderstanding, or incomplete state responses. This information was used to refine materials to be sent to the states. HumRRO prepared a CD for each state containing the following information already collected:

- Test characteristics file
- State assessment scale score data file
- State assessment percentages proficient data file
- Data verification checklist, which state officials were asked to sign
- Two-page summary of basic trends
- Directions for completing the verification, tailored for each state

State verification mailings were sent from November 29, 2007, through December 28, 2007, with staggered due dates between January 4 and January 24, 2008.

CEP and HumRRO took the following steps to ensure that officials from every state verified the accuracy of their state's data and filled in missing information.

1. **Initial letter.** In November/December 2007, CEP President Jack Jennings sent a letter and a CD to three people in each state: the state superintendent or other chief state school officer, the deputy superintendent, and the assessment director. The letter asked state officials to verify the data on the CD and make necessary corrections or additions by a specified date in January 2008. The letter included an offer from CEP to help defray costs. Also included in the packet was a printed copy of the summary of basic trends.

2. **Follow-up efforts.** HumRRO staff and CEP's president communicated with individual states to maximize participation. The last data file was received April 4, 2008. The last bit of related documentation (the test characteristics file and verification checklist) was received April 9, 2008.

3. **Questioning anomalies.** During the analysis phase, HumRRO staff contacted state officials when anomalies in the data arose, up until May 20, 2008.

During the verification process, some state officials added missing data themselves. Other states hired outside help (sometimes with CEP funding) to add missing data. Still others provided raw data in various forms to HumRRO staff, who added the data to the appropriate tables and sent the revised file back to the state.

Most states returned modified data Excel files and modified files of test characteristics file by e-mail and faxed the signed verification checklist. Throughout the verification period, HumRRO maintained a log of all e-mail and phone communications with states (260 records).

After each state verified its data, HumRRO staff produced a draft "profile," consisting of data tables and figures for that state. A HumRRO staff person was assigned to conduct a series of quality control checks on each state profile, primarily to confirm that the profile matched the original source data file.

Analyzing Data

Analyses of the data collected were conducted from January through May 2008. Data were analyzed using a consistent set of rules for all states.

SELECTION OF YEARS FOR ANALYSIS

The data collected for last year's achievement study ended in 2006, at the latest. In this year's study, the trend lines for the post-NCLB period were extended to 2007 for nearly all states.[1] These 2007 data come from tests administered in spring 2007 or fall 2006 in states with fall testing. (Results from state tests administered during school year 2007-08 typically are not available until later in the summer of 2008.)

[1] The test score trends for this report do not include any data for Hawaii and Kentucky. Although 2007 data are available from these states, both states instituted new tests in 2007, and the 2007 results are not comparable to those from previous assessments. Reporting 2007 data alone would not have enabled us to report any three-year trends or even two years of changes in test results. Other states were missing 2007 data at some grades. In addition, Oregon was missing 2007 percentage proficient data but had the scale scores needed to compute effect sizes.

This year we did not analyze national trends for the years preceding NCLB, since the pre-NCLB trends we reported last year would not have changed. Moreover, the number of states with comparable pre- and post-NCLB data, which was already limited, is likely to decrease over time as states make further changes in their testing systems. In states where pre-NCLB results remain comparable with post-NCLB results, we have included the pre-NCLB data in the state profiles that accompany this report.

Where comparable data were available, we analyzed trends for the entire period from 2002 through 2007. Where data could not be obtained for the entire period or were not comparable, we used other starting or ending years based on available data. If a new test was introduced in 2006, we showed data from 2006 and 2007 in the state profiles and referred to any movement in achievement over the two years as a change rather than a trend. However, states with just two years of comparable data were not included in the national summary of trends discussed in this report. When a new test was introduced in 2007, we used data from 2002 through 2006 or whatever period provided three years of comparable data, rather than showing a single year of data from 2007, which would not meaningful.

In seven states, the years of available data were different for percentages proficient and effect sizes. In the cases where the years were different and the two indicators produced contradictory trends, we have noted this difference in the tables in chapters 3 and 4.

SELECTION OF GRADES TO REPORT AND ANALYZE

In each state's data profile, the percentages of students scoring at the proficient level or above for all tested grades were presented in tables. However, for graphs of percentages proficient, all effect size analyses, and all achievement gap analyses, a representative grade was identified at each grade span (elementary, middle, high school). HumRRO reviewed data for all the grades tested between 2002 and 2007 to make decisions about which specific grades to report on and analyze. These decisions were based on a fixed set of criteria applied consistently across all states and developed without regard to whether achievement was high or low at a given grade.

Preferred grades were grade 4 to represent the elementary span, grade 8 in the middle span, and grade 10 in high school. If tests in these grades were comparable for a three-year period and thus supported trend analysis, these grades were selected. If these grades did not have comparable tests for the minimum period (or if the state tested a different high school grade for NCLB purposes), adjacent grades were reviewed and an alternate was selected if it had a longer trend line. The representative grade was selected without regard to whether performance was high or low at that grade or what pattern of performance trends appeared over time.

In most states, the analysis of the overall percentages of students scoring proficient in the state profile included more grades than the analyses of achievement gaps or effect sizes, which focused on a representative grade at each grade span. As a result, data are available for grade levels and years that are not analyzed in this report. These raw data have been posted on CEP's Web site, www.cep-dc.org.

PROCESS FOR ANALYZING STATE DATA

HumRRO provided the draft profiles to CEP as they were completed. As anomalies were encountered, HumRRO and CEP staff conferred to expand or clarify the rules for analysis and reporting.

A senior CEP consultant with expertise in educational testing and a CEP research associate analyzed the data in each state's profile to determine both overall trends and achievement gap trends. For all of the achievement analyses, HumRRO and CEP looked separately at the elementary, middle, and high school levels. As a cross check, the HumRRO senior project member did an independent analysis of the main trends for each state. Where the results were different, the three analysts reviewed state test data until they agreed an accurate finding had been reached.

The CEP consultant who did the trends analysis also wrote narrative findings summarizing the trends identified for each state, while the CEP research associate developed a state-by-state national summary table of trends. Another senior CEP consultant developed templates for tables and edited the state profiles, including the tables and narrative findings, to help present the information in the state profiles as clearly as possible.

BIAS CHECK FOR NAEP RESULTS

This year we compared trends on each state's test to trends on NAEP for the same state. To determine whether the unweighted analysis of NAEP results used in these comparisons were biased, we compared the average of the state-by-state NAEP data to the national NAEP data from the main NAEP (not the NAEP "long-term trend" assessment). As table 7-A indicates, national NAEP results were quite similar to the average of the state NAEP results for both the percentages of students performing at or above the basic level and effect sizes. These findings suggest results from unweighted state NAEP performance data can be generalized to the nation as a whole.

Table 7-A. Average of State NAEP Results and National NAEP Results

	Percentage Basic or Above		Mean Scale Score	
	State Average	National	State Average	National
Grade 4 reading	67%	66%	221	220
Grade 8 reading	74%	73%	263	261
Grade 4 math	82%	81%	240	239
Grade 8 math	72%	70%	281	280

Appendix 2. More about Standard Deviations and Relationship of Gaps to Cut Scores

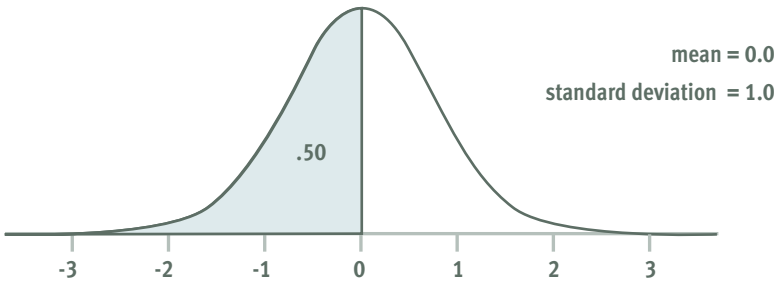
Effect size, an indicator of achievement used in this report, is based on mean, or average, scale scores on a test and a statistic called a standard deviation. Using graphs, this appendix explains in more detail what standard deviations are. The appendix also presents graphs to illustrate a concept noted throughout this report—namely, that the location of the cut score on a test makes a difference in the apparent size of an achievement gap between two subgroups.

What Are Standard Deviations?

Normal curve figures, such as the ones below from a statistics text by David Stockburger (1998), are used to graphically represent the distribution of scores on any administration of a test. The largest numbers of test-takers' scores cluster close to the middle or high point of the curve, while fewer scores are situated at the very low and very high ends.

Three areas on a standard normal curve are useful for interpreting test scores. The first is at point 0, which is the mean, or average, test score. As shown in **figure 7-A**, 50% percent of the scores on the test are below the mean and 50% are above.

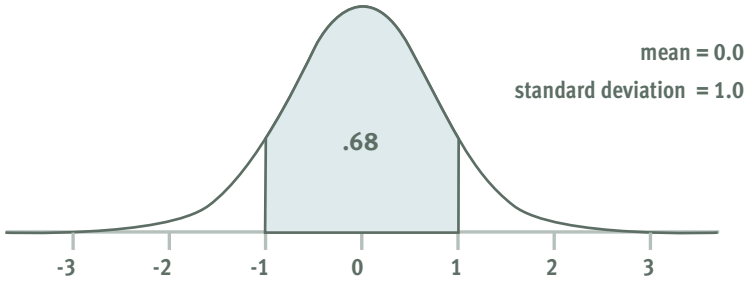
Figure 7-A. The Mean



Source: Stockburger, 1998.

The second area, shaded in **figure 7-B**, is within +1 or -1 standard deviation from the mean; 68% of the scores on a given test fall within this area. One standard deviation above the mean captures 34% of scores (half of 68%).

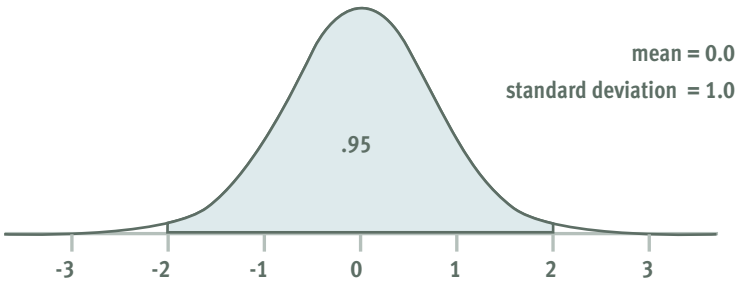
Figure 7-B. One Standard Deviation Above and Below the Mean



Source: Stockburger, 1998.

The third area of interest, the shaded area in **figure 7-C**, is between +2 or -2 standard deviations. This area accounts for 95% of the scores on a given test.

Figure 7-C. Two Standard Deviations Above and Below the Mean



Source: Stockburger, 1998.

Let's say a test is scored on a scale from 1 to 1000, and the mean score is 500 and the standard deviation is 80. This means that the scores of 68% of test-takers are between 420 (500-80) and 580 (500 + 80). Similarly, 95% of the scores would fall between 340 and 660.

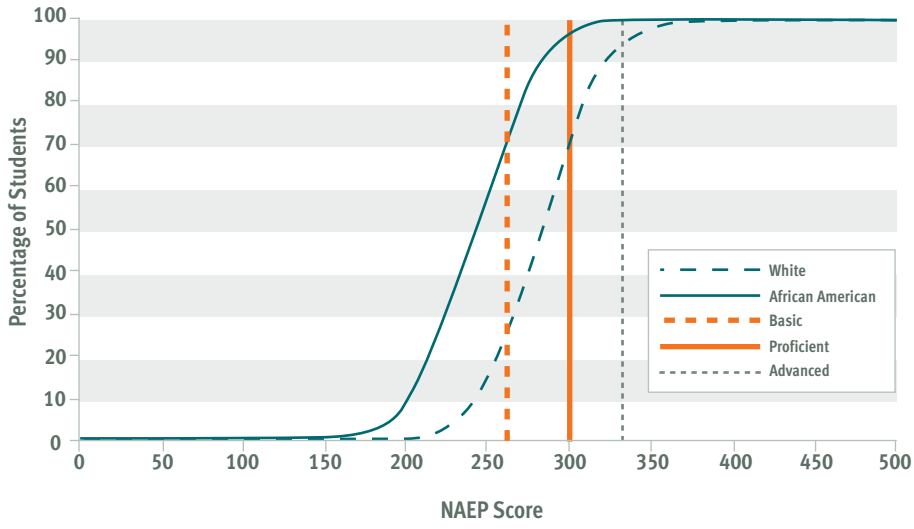
Since the percentage of test-takers who score within one or two standard deviations of the mean is always the same for any test, the standard deviation is a common unit of measurement that can be used to make limited comparisons of groups of test-takers. In this study, effect sizes are used, which is the proportion of the difference between two years of test data or two subgroups of students in standard deviation units.

Cut Scores and Gaps

Changes in instruction can affect the size of achievement gaps, but so can other factors. An important issue to consider when looking at achievement gaps is the location of the cut score students must reach to be considered "proficient" on a test. Research shows that where the proficiency cut score is set makes a difference in the apparent size of the gap. If a proficiency cut score is very high or low, so that almost everyone reaches the cut score or almost nobody reaches it, there is little apparent gap. A cut score closer to the mean test score will be more sensitive in detecting achievement gaps, and the gaps between subgroups will appear larger.

This was illustrated graphically by Paul Holland (2002). **Figure 7-D** shows the results of the 2000 administration of the math portion of the National Assessment of Educational Progress for 8th grade African American and white students. The test was scored on a scale of 0-500, with the cut score for the basic level of achievement set at 262, proficient at 299, and advanced at 333. The figure shows the percentage of students in each group that scored *at or below* a certain level on NAEP. The x axis is the score, and the y axis is the percentage of students achieving that score or scoring below it. So, about 25% of white students scored at or below 262 (basic)—marked with a dashed vertical line in figure 7-D—while 75% exceeded this score. About 70% of African American students scored at or below 262, while about 30% exceeded this score. Therefore, at the basic level, the achievement gap between African American and white students is about 45 percentage points—quite large.

Figure 7-D. African American-White Achievement Gap, NAEP Mathematics 2008 Grade 8



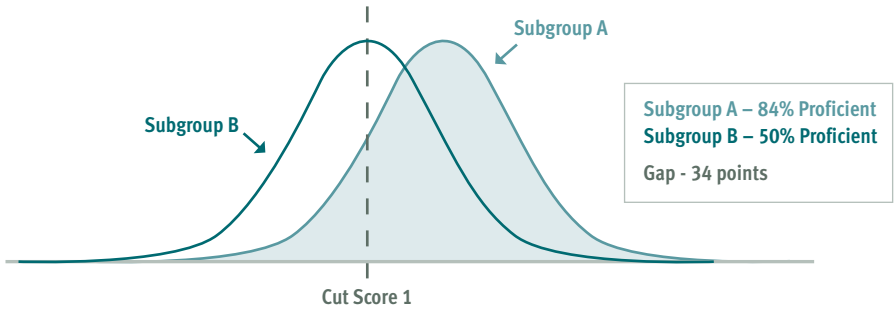
Source: Holland, 2002.

However, the achievement gap picture changes as one moves along the score scale. At the proficient level of 299—marked with a solid vertical line in figure 7-D—the black-white gap shrinks to about 30 percentage points. As one moves toward the advanced cut score of 333 (shown in the figure as a dotted vertical line), the gap continues to shrink until it reaches about 6 percentage points at the advanced level. The same is true at the low end of the scale, where the gap is also a lot smaller.

As this NAEP example shows, choosing a cut score of 262, 299, or 333 will have a dramatic impact on the apparent size of the achievement gap between African American and white students. The gap is larger at the middle of the NAEP score scale than at the extremes.

Figure 7-E illustrates this phenomenon in another way. The figure consists of two normal distributions of test scores for two subgroups of students, subgroup A and subgroup B. The figure displays a hypothetical example whereby the initial cut score (cut score 1) is set so that 84% of the students in subgroup A score *at or above* the cut score, compared with 50% of the students in subgroup B. (The areas to the right of the cut score under both curves represent the students who pass.) Therefore, the gap in percentages proficient between the two groups is 34 percentage points.

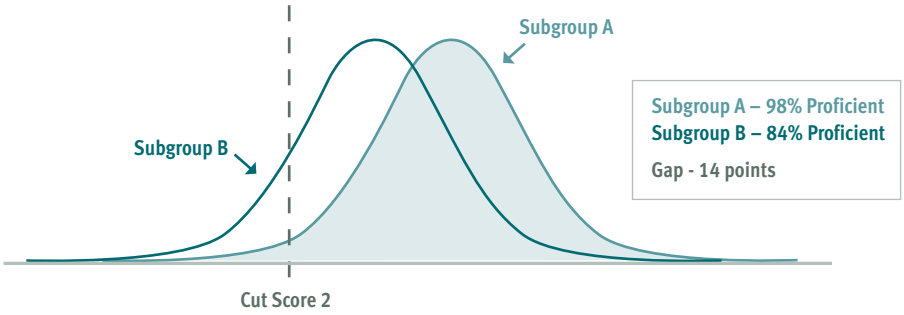
Figure 7-E. Size of Gaps in Percentages Proficient with a Cut Score at the Mean



Source: Center on Education Policy.

If a state were to set an easier cut score, represented by cut score 2 in **figure 7-F**, more students would meet or exceed it. At that point, 98% of subgroup A students and 84% of subgroup B students would pass, and the achievement gap would be reduced to 14 percentage points.

Figure 7-F. Size of Gaps in Percentages Proficient with a Lower Cut Score



Source: Center on Education Policy.

Therefore, anyone examining trends in achievement gaps must take into account the location of the proficiency cut score, as well as possible changes in cut scores.

Credits and Acknowledgments

This report was written by Nancy Kober, Naomi Chudowsky, and Victor Chudowsky, CEP consultants. Naomi and Victor Chudowsky also analyzed the state test data for this study. A team from the Human Resources Research Organization (HumRRO) under the direction of Sunny Becker conducted extensive research and analysis for the report and wrote papers that formed the basis of the NAEP chapters. Diane Stark Rentner, CEP's director of national programs, oversaw the study project for CEP, communicated with the expert panel, and provided advice and assistance for all aspects of the study, including the state profiles. Jennifer McMurrer, CEP's research associate, assisted with data analysis, table development, and the project in general. Jack Jennings, CEP's president and CEO, provided advice and assistance. Other key HumRRO staff involved in the project were Hilary Campbell, Monica Gribben, and Lauress Wise.

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

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

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

www.cep-dc.org

Center on Education Policy
1001 Connecticut Avenue, NW, Suite 522
Washington, D.C. 20036

tel: 202.822.8065
fax: 202.822.6008

e: cep-dc@cep-dc.org
w: www.cep-dc.org