Policy Information Report

Windows on Achievement and Inequality

ETS®

Listening. Learning. Leading.®

This report was written by:

Paul E. Barton
Richard J. Coley

Additional copies of this report can be ordered for $15 (prepaid) from:

Policy Information Center
Mail Stop 19-R
Educational Testing Service
Rosedale Road
Princeton, NJ 08541-0001
609-734-5949
pic@ets.org

Copies can be downloaded from:
www.ets.org/research/pic

April 2008
Policy Evaluation and
Research Center
Policy Information Center
Educational Testing Service

Table of Contents

Preface

In this ambitious report, Paul Barton and Richard Coley take us beyond typical data and information about the status of educational achievement in the United States and about gaps in achievement among the nation's students. They shift the focus and then take us on an exploration of data that address some often-neglected questions.

To start, they explore how a child's development is affected by parent-child interactions during the child's earliest years of life. Then they look at children in kindergarten and provide evidence of an already burgeoning gap.

As required by the No Child Left Behind Act (NCLB), educators are continuously monitoring whether more or fewer students are scoring at a level termed *Proficient*. But what about changes in the distribution of scores? Barton and Coley examine what is happening to both top-performing and lower-performing students and how the distribution of scores is changing in the United States.

Traditionally, typical reports on educational achievement focus on how much students know about math, for example, at the end of a school year. But the public desires more information about how much students learned during the school year. Barton and Coley make a case for measuring such growth.

NCLB also requires that states provide information about gaps in student performance among racial/ethnic groups. Barton and Coley warn us about the pitfalls of comparing such numbers across states, pointing out that different states have established different definitions of what it means to be *Proficient*. And even when common ground can be found, they say, test scores are often conveyed as abstract numbers with vague meanings. The authors demystify these numbers by providing examples of the kinds of knowledge and skills that students are likely to be able to demonstrate at particular score levels.

Frequently missing from the public's view of achievement are the school and life conditions that may influence student performance in school. The authors lead us to several windows that allow us to see how changes in the demographic characteristics of the student population over the past several decades have affected national test scores.

And last, but certainly not least in an increasingly competitive world, Barton and Coley seek to provide a simple, summative view of where U.S. students rank globally. They do this by summarizing results from international assessments, which vary by participating countries, subjects assessed, and grades and ages covered.

There are, the authors conclude, many windows in the house of achievement that parents, educators, policymakers, researchers, and the media should be looking through — many more than are now open. By pulling the blinds on a few of these windows, Barton and Coley illuminate aspects of education and achievement that warrant further attention.

Michael T. Nettles
Senior Vice President
Policy Evaluation and Research Center

Acknowledgments

The authors appreciate and acknowledge the thoughtful feedback, comments, criticism, and suggestions made by the following reviewers of the draft report: Henry Braun and Drew Gitomer of Educational Testing Service and Margaret E. Goertz of the Center for Policy Research in Education at the University of Pennsylvania. The authors are also grateful for statistical and data analysis support from several ETS colleagues. The report was edited by Janet Levy. Marita Gray designed the cover and Christina Guzikowski provided desktop publishing. Bill Petzinger and Jodi Krasna coordinated production. Any errors of fact or interpretation are those of the authors.

For people with a deep interest in the success of the education enterprise and in eliminating achievement gaps, the limited perspectives offered in newspapers and on television will not suffice. To reach — or reach for — a deeper understanding, the facts must be viewed through all the available windows in the schoolhouse. This section highlights some of the data that are explored in the full report.

In the Nursery. Learning, and developing the ability to learn and think, begins in the nursery. The first few years are critical, as researchers Betty Hart and Todd Risley demonstrated by closely observing the interactions of parents and their babies from birth through age 3, and extrapolating the results to age 4.

By age 4, the average child in a professional family in the study heard:

- about 20 million more words than the average child in a working-class family, and
- about 35 million more words than the average child in a welfare family.

Growth in the children's vocabularies paralleled the quantity of words they heard from their parents. So, by this young age, the vocabulary of the average child in the professional families was larger than that of the average *parent* in the welfare families. Well before any formal public schooling, children are vastly different in their achievements.

How much effort would it take to equalize the language experience of the children in welfare families with the children in working-class families? Hart and Risley estimate that it would take 41 hours per week of "out-of-home experience as rich in words addressed to the children as that in the average professional home."

In the Sandbox. This report focuses on the children again once they reach kindergarten. Between age 3 and kindergarten, children have highly different degrees of learning — and achievement — opportunities in their families, among care givers, and — if they are fortunate — in pre-school educational settings. The question we pursue is, how are the children doing as they begin their formal education?

According to a groundbreaking national longitudinal study by the National Center for Education Statistics (NCES), 65 percent of children entering kindergarten could recognize the letters of the alphabet. There was considerable variation by race/ethnicity and socioeconomic level of the parents.

Thirty percent of kindergartners were able to understand the letter-sound relationship at the beginning of words. Again there was variation among racial/ethnic groups:

- Asian American students, 44 percent
- White students, 34 percent
- Black students, 20 percent
- Hispanic students, 20 percent

Similar differences were found in other aspects of reading, and in mathematics. These students were studied again by NCES in the first grade and are being tracked through fifth grade. We check on them again in fourth grade, when the National Assessment of Educational Progress (NAEP) commences its assessments.

Standard Measures of Student Proficiency. With the way achievement is now reported in state testing systems, as required under the federal No Child Left Behind Act and as measured by NAEP, standard practice has become to look at the percent of students reaching or exceeding a test score or level labeled *Proficient*. In the parlance of the educational testing community, this is a "cut point." This report provides considerable information on trends in the percentage of students reaching NAEP's three achievement levels — *Basic*, *Proficient*, and *Advanced*.

In fourth-grade reading in 2007, 43 percent of White students scored at the *Proficient* level, compared with 17 percent and 14 percent, respectively, of Hispanic and Black students. This was an improvement from the 1990s for White, Hispanic, and Black students.

Drawing on NAEP, this report provides such national information for fourth- and eighth-grade reading and for fourth- and eighth-grade mathematics, where the news was considerably better.

State Testing. Cut points are used almost exclusively to report the educational progress of states to the public. These measurements form a basis for school accountability and for sanctions against schools that are judged to be performing poorly. They are also now the basis for reporting gaps in achievement by race/ethnicity, perhaps the most important feature of No Child Left Behind. As critical as it may be to measure achievement in terms of reaching a set standard, going beyond this measure provides a broader view of achievement gaps.

- In terms of trends, cut points tell us only about the movement of a relatively few students near the cut

point, whether above or below it. But we need to know about the other students, as well.

- When looking at the difference in performance of racial and ethnic subgroups, the size of the achievement gap (in terms of the percent reaching or exceeding the cut point) will vary, depending on how high the cut point is set. At a low cut point, all groups may reach or exceed it, so the gap appears to be small. Since states use different tests and different cut points, it is impossible to compare states on the basis of achievement gaps on state tests.
- State comparisons of achievement gaps among racial/ethnic groups can be made by examining the percentage reaching the *Proficient* level, as reported by NAEP; this report provides these comparisons. The data show that *where* the cut point is placed on the scale makes a difference in the size of the gap.

Reporting by Averages. While reporting trends of those who reach a score cut point involves relatively few students, reporting by averages considers *all* student scores. Comparing states by average NAEP scores is a way to capture the size of and changes in achievement gaps, based on the scores of all of the states' students.

In the subject of reading, using NAEP Long-Term Trend data, the best news is for 9-year-olds, who performed higher, on average, in 2004 than in any previous year. This gain was shared by White, Black, and Hispanic 9-year-olds. For 13- and 17-year-olds, the picture is cloudier. In mathematics, average scores increased in 2004 for 9- and 13-year-olds in all racial/ethnic groups.

Reporting Scores Up and Down the Scale: The Score Distribution. Having looked at the percentage of students reaching a cut point on a scale and the average scores of all students, this report shares research that may be more demanding of the reader's attention but that captures much richer information about student performance.

A more panoramic view shows student performance by percentiles up and down the score scale. This report compares scores at the 90th percentile — that is, the score at which 90 percent of students score below — as well as the 75th, 50th, 25th, and 10th percentiles. This analysis is provided in reading and mathematics, at two points in time, for students at ages 9, 13, and 17.

For all 9-year-olds, there were score increases in reading between 1990 and 2004 at the 50th, 25th, and 10th percentiles, showing that the improvement was at the middle and lower half of the score distribution. There was no improvement for students in the top of the score distribution. Black 9-year-olds, though, gained at all percentile levels, and Hispanic 9-year-olds gained at all levels except at the 90th percentile. The good news for reading was not carried over to 13- and 17-year-olds; however, improvement was widespread in mathematics.

This report also provides a trend analysis of national data that groups students into quartiles (or fourths) — the average of the top fourth of students, and so on. Examining quartiles helps determine whether the goal set by the Education Summit of Governors in 1989 has been met: that "the academic performance of all students … will increase significantly in every quartile, and the distribution of minority students in each quartile will more closely reflect the student population as a whole."

By examining quartiles, we can track score changes at different parts of the score distribution. For example, we can see how the average score of students in each quartile has changed and also track changes in the achievement gap in each quartile. For reading, we are able to look at the period from 1975 to 2004; for math, from 1978 to 2004.

For 17-year-olds, we see a big change in reading among minority students. Scores jumped at each quartile between 1975 and 1990 but have not improved since, and have fallen in the bottom quartile since 1990.

Using a different set of NAEP data, an analysis by quartiles was extended to the states to track the performance of the top and bottom quartiles. And for comparison, this report also shows the percentage of students reaching the NAEP *Proficient* level, as well as average scores. The tabulation of states compares them on the basis of whether they improved, stayed the same, or did worse on eighth-grade reading (from 2002 to 2007) and math (from 2000 to 2007). Looking at eighth-grade reading, for example, we see contrasts, depending on which performance measure is used:

- While none of the states improved in average score or the percent of students scoring at or above the *Proficient* level, five increased in the score of the top quartile and five increased in the score of the bottom quartile.
- While most states showed no change on the four measures, 12 showed a decline in the average score, three showed a decline in the percent *Proficient*, 11 declined in the top quartile score, and 14 declined in the bottom quartile score.

These examples and the differences they reveal illustrate why more than standard measures are needed to develop a more complete picture of change.

Measures of Student Learning in the Classroom. It may come as a surprise, after all of the views provided in this report, to be warned that *none* of the measures presented consider what individual students, classrooms of students, or whole schools of students learned over the course of a year's instruction. The measures discussed thus far compare the scores of students at, say, the end of eighth grade with the scores of *different* students from prior years at the end of eighth grade. There are only a few places in the United States where the *gain* in what students know is measured. This approach is referred to as measuring gain, growth, or value added, and provides an important perspective on student achievement and gaps. Schools should be held accountable exclusively for what students learned in the classroom — not for out-of-school experiences that affected students.

While NAEP does not report how much students grow in knowledge from the fourth to the eighth grade, for example, the fact that the assessments have been given four years apart and the scores reported on a common scale enables us to use the data to estimate how much students' knowledge has grown over the four years. This report tracks reading results among a cohort of fourth-graders in 1994 and eighth-graders in 1998, and tracks math results among a cohort of students who were fourth-graders in 1996 and eighth-graders in 2000. We report how much these scores increased on the 0 to 500 NAEP scale over the four-year period.

While the differences among various racial/ethnic groups in achievement are large when their end-of-year scores are compared, the difference in how much they grew in knowledge between the fourth and eighth grades is small. In reading, the average growth score was 50 points. By subgroups, growth was as follows:

- 56 points for Black students
- 54 points for Hispanic students
- 48 points for White students
- 47 points for American Indian students
- 42 points for Asian/Pacific Islander students

Math was different, with Black and Hispanic students trailing White and Asian/Pacific Islander students. A prior analysis of the same type for 1992-96 showed all subgroups growing by about the same amount.

White and minority students enter school with quite different levels of knowledge, but they increase their knowledge by similar amounts in the classroom. The result is that the gap between them remains about the same. This finding is consistent with statistics that show students' levels of growth from the fourth to eighth grade.

What difference does using a growth measure make in the rankings of performance among the states? A lot. For example, in reading, Maine was at the top in the level of knowledge (as measured by the average score), but dropped to fourth-from-the-bottom in terms of score growth from fourth to eighth grade.

Exactly What Can Students Do? NAEP reports achievement by showing scores along scales it has created. But as indispensable as they are, scale scores are limited in their helpfulness, for they convey very abstract ideas.

But NAEP also makes it possible to examine the specific kinds of questions students are able to answer and the problems they can solve at various points along the NAEP scale — often referred to as an "item map."

On the eighth-grade reading map, the average Asian/Pacific Islander and White student scores at a point along the scale where they likely "can use task directions and prior knowledge to make a comparison," and the average Black student "can locate specific information in a detailed document."

In mathematics at grade eight, the average White student "can solve problems of square root," while the average Black student "can draw the reflection of a figure."

It would likely improve communications with teachers, parents, students, and the public generally if *all* achievement tests were reported in this way, as well as in standard ways, to communicate what students can and cannot do.

Deconstructing Achievement Gaps. After considering different ways of looking at achievement gaps, the questions became: What is behind these gaps? How did they originate? How are they sustained?

These questions are answered by synthesizing volumes of research on the correlates of achievement — efforts to identify the life and school conditions and experiences associated with cognition and school achievement. Fourteen such correlates have been identified: eight associated with life before and after school, and six associated with life in school.

The next step was to see what the gaps were by income and by race/ethnicity in these critical conditions and experiences. To illustrate:

- An out-of-school correlate of achievement is reading to young children; minority children, on average, are read to considerably less frequently than are other children.
- An in-school correlate is teacher experience; minority children, on average, are taught by less experienced teachers than are other children.

This sort of information helps provide a road map to the elimination of inequality among different racial/ethnic populations.

Looking at Changing Achievement and Demographics Together. Research shows that Black and Hispanic students, on average, have lower achievement scores than White and Asian American students, and also that the composition of the total population is changing over time. The net result, in terms of average achievement levels for all children combined, has never been clear.

What would the average achievement level for the United States be now, if the racial/ethnic populations were proportionally similar to what they were in the 1970s? The possibly surprising answer is that they would not be very different.

Understanding International Studies of Achievement. Another form of inequality in educational achievement is seen in international comparison studies. These studies receive a lot of publicity, particularly if the United States does not fare well, as is often the case in mathematics. The studies receive less press if the United States does well, as it usually does in reading. Regardless, the results are hard to track, with assessments made at different times and in different countries, with students of different ages, and on different subjects.

To help understand these assessments and surveys, two researchers have simply added them altogether, and then looked at the U.S. standing in the composite.

- In *reading*, 13 percent of the participating nations scored above the United States, 44 percent had scores that were equivalent, and 44 percent scored below.
- In *mathematics*, 53 percent scored above the United States, 32 percent had equivalent scores, and 15 percent scored below.
- In *science*, 35 percent scored above the United States, 40 percent had equivalent scores, and 25 percent scored below.
- In *civics*, none scored above the United States, 33 percent had equivalent scores, and 67 percent scored below.
- In aggregating *overall*, 24 percent scored above the United States, 37 percent had equivalent scores, and 35 percent are below.

These results can be interpreted differently, depending on what one's expectations are for the United States and where one thinks the country needs to be. The data in this report on educational inequality in the United States are highly relevant to this discussion because they inform about the United States' international ranking.

* *

Few things are more important to the United States' economic well-being and strength internationally, and to the strength of its democracy, than the education of its citizens. That said, for individuals, little is more important than equality in educational opportunities and, as a result, achievement.

To reach a state of greater equality, we must first see clearly through all the windows from which we can view educational achievement. This will enable us to accurately judge the success of U.S. schools, as well as U.S. success overall and by state and community.

As important as improving the overall education system and reducing achievement gaps have become, educators, policy officials, and the general public have been provided with only limited views of student performance. The No Child Left Behind Act makes significant progress in expanding these views. Important aspects of the law require that all the states participate in the National Assessment of Educational Progress (NAEP), that test scores for subgroups of students be collected and publicized, and that students with disabilities and English-language learners are included in the assessments. The purpose of this report is to expand the view into a more complete and better-developed picture of student achievement. To see into the whole house, the shades have to be up on every window.

There are, of course, thick, official documents, such as the comprehensive *Digest of Education Statistics* and the *Condition of Education* published by the National Center for Education Statistics (NCES) and many large cross-sectional and longitudinal databases that contain a plethora of important information on educational achievement and its correlates. But most people rely on the education information and interpretations that find their way into reading materials in their mailboxes or office in-boxes, or on television or the Internet. But what is typically found in such media are a few well-known statistics that leave a lot of important questions unanswered:

- While the media reports annual trends in college-admissions scores, what is the whole population of students learning in school and how is this changing?
- Reports from the well-respected NAEP provide much detail on average student achievement and progress and identify the proportion of students that reach different achievement levels like *Basic* and *Proficient*. But to expand our understanding of student achievement, we need to examine the entire distribution of test scores that are provided by NAEP.
- While the names of schools designated as failing or "in need of improvement" are published in the newspaper, we lack a complete picture of achievement gaps that may exist in schools that manage to meet the strict statistical requirements of the law.
- News flashes focus on various international assessments that are conducted in different grades or at different ages, in different subjects, and in different sets of countries. But what is the net of it all, how does U.S. performance stack up when all of the assessments are considered, and how much does the large achievement gap in the United States affect the country's international standing?

Each time we receive more information, our understanding of student achievement and progress improves. In this report, we try to look at achievement from its many sides, beginning even before formal schooling starts.

Specifically, the report is about:

- Understanding how cognition and vocabulary develop in the first three years of life.
- Getting beyond the cut points and the averages typically reported on tests, to identify and understand the performance of students at different points on the score distribution, and determining how these scoring gaps have changed over time.
- Expanding our understanding of student achievement beyond the total knowledge students have at the end of a school year to determine how student achievements *grow* while students are *in* school, and how gaps in growth compare with gaps in total knowledge.
- Getting beyond the abstraction of "scale scores" to see what specific tasks students can perform, and if students' performances differ by race/ethnicity.
- Acknowledging the large demographic changes that have taken place in past decades (and that continue to take place) and exploring what national average achievement scores might look like if there had not been such changes. And, relatedly, identifying early life and school factors that are strongly related to school achievement, and describing the gaps that exist in these critical experiences and conditions.
- Making sense of different international assessments to understand how the United States compares with other developed nations.

This report is not a compendium or statistical abstract — it is only 46 pages. Rather, it tries to illuminate all rooms in the achievement house by opening as many windows as the data allow. While we recognize that this expanded view still provides only a glimpse of the total picture, we hope to improve insight and understanding. We also wish to convey that those who prepare public reports on education or who write about student achievement need to provide a fuller picture from all the data they collect.

The first window we encounter peers into the nursery. This is where achievement gaps begin. While there have been many studies about what happens in the early years of life and how early experiences affect cognition and language acquisition, none has been as thorough as the Betty Hart and Todd Risley study of children in functional families from birth to age 3.[1] Recording and monitoring many aspects of parent-child interactions over the course of 36 months, the researchers noted the children's progress.

Like Their Parents. Hart and Risley found that in vocabulary, language, and interaction styles, the children begin to mimic their parents: "When we listened to the children, we seemed to hear their parents speaking; when we watched the children play at parenting their dolls, we seemed to see the future of their own children."[2]

Hart and Risley listened not just to the words parents used, but also to the tone exhibited by family members. The researchers then observed what the children said and how they spoke at age 3.

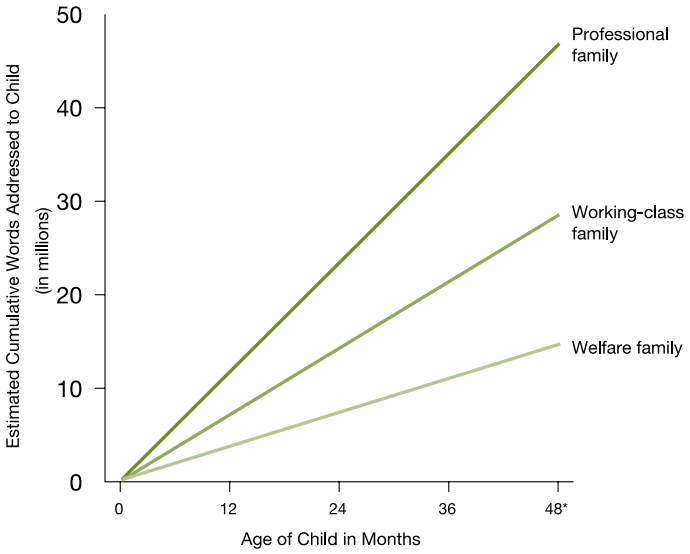
An example of what they observed was that, in working-class families, "about half of all feedback was affirmative among family members when the children were 13-18 months old; similarly, about half the feedback given by the child at 35-36 months was affirmative."[3]

An affirmative tone was slightly more prevalent in professional families, and the children shared this. However, in the families on welfare, about 80 percent of parents' feedback was negative; similarly, when the children reached age 3, almost 80 percent of their feedback to family members was negative. Hart and Risley reported that there was "a consistent and pervasive negative feedback tone. The general exchanges among family members — parents and older siblings — were negative." In the families on welfare, the researchers generally found a "poverty of experience being transmitted across generations." And, they said, "We could see why a few hours of intensive intervention at age 4 had so little impact on the magnitude of the difference in communicative experience that resulted from those first three years."[4]

A summary of the researchers' findings related to language exchanges is provided in Figure 1.

Figure 1

Estimated Cumulative Differences in Language Experience by 4 Years of Age



*Projected from 36 to 48 months
Source: Hart and Risley, 1995.

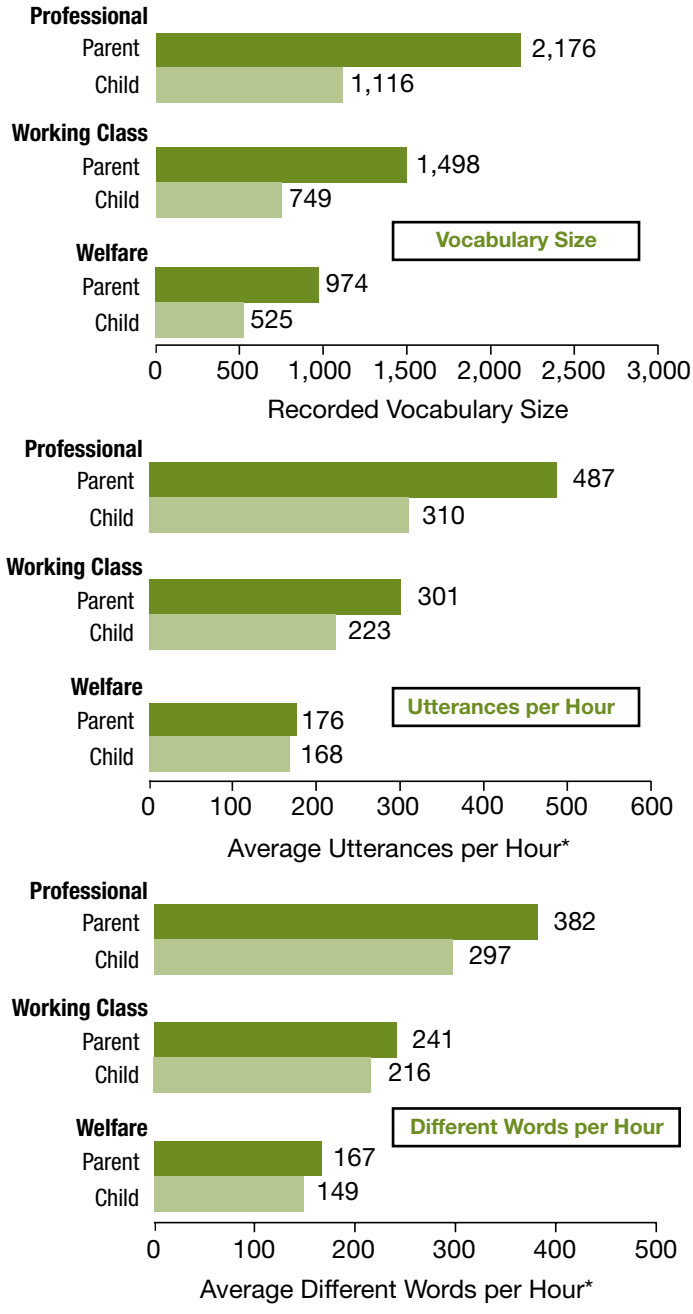
Figure 2 illustrates associations between parents and children in language use. For both parents and children, larger vocabularies and frequency of speaking are highest in professional families, next highest in working-class families, and lowest in welfare families. More striking is that the vocabulary of children in professional families was greater than that of parents in welfare families.

Parent-Child Interaction. By what process did the observed children come to mirror their parents? A key observation was the frequency of words addressed to the baby and child: Figure 1 shows the estimated number of words addressed to the children over 36 months, with the trend extrapolated through 48 months.

[1] Betty Hart and Todd R. Risley, *Meaningful Differences in the Everyday Experience of Young American Children*, Paul H. Brookes Publishing Co., 1995.
[2] Hart and Risley, 1995, p. 176.
[3] Hart and Risley, 1995, p. 177.
[4] Hart and Risley, 1995, p. 180.

Figure 2

Measures of Parent and Child Language, by Socioeconomic Status



*Child utterances and different words were averaged over ages 13 to 36 months.
Source: Hart and Risley, 1995.

The differences are huge among the professional, working-class, and welfare families. By the end of four years, the average child in a professional family hears about 20 million more words than do children in a working-class family, and about 35 million more than the children in welfare families.

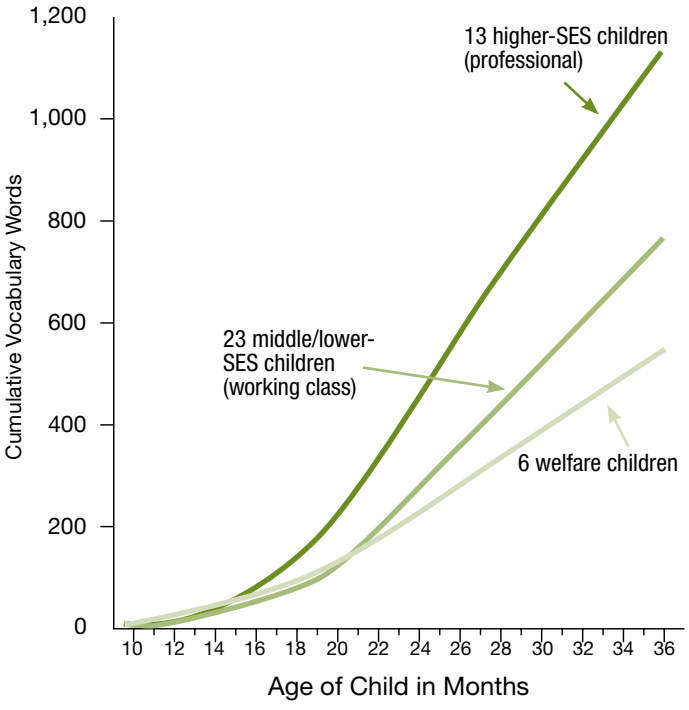
How does exposure to words affect the children's vocabularies? See Figure 3 for insight. The lines begin to diverge between children in professional families and the others at around 15 months, when children start to talk. The divergence between children from working-class and welfare families begins after about 22 months. By 36 months, the vocabulary of children in professional families is more than double that of children in welfare families.

Closing the Gap. Hart and Risley ask: "Is it possible to change children's lives in a generation?" They describe the effort that would be required with the following:

> To ensure that an average welfare child had a weekly amount of experience equal that of the average child in a working-class family, merely in terms of hours of language experience of any kind (words heard), 41 hours per week of out-of-home experience as rich in words addressed to the children as that in an average professional home would be required . . . welfare children would need to be in substantial care 40 hours every week from birth onward.[5]

Figure 3

Cumulative Vocabulary Recorded, by Socioeconomic Status and Age



Source: Hart and Risley, 1995.

[5] Hart and Risley, 1995, p. 205.

With the right determination, the researchers say, closing this gap is possible. They cite an early intervention effort, the Milwaukee Project, through which infants with mothers whose IQs were 75 or below were enrolled at 6 to 8 weeks of age in out-of-home, full-day care. By age 8, the children were equal to the national average in accomplishments.[6] But this level of investment is beyond any that has been seriously proposed, or perhaps ever imagined, by policymakers. Given the size of the gaps in early development, depending on differences in parent-child interactions, and the resources and time needed to equalize development from outside interventions, it is not surprising that we see striking inequalities in other windows of the achievement house.

[6] There was a transition period when the caregiver was in the home participating in the parenting.

This report now fast forwards from the nursery to kindergarten, where we observe children's readiness to start learning in school, while mindful of the developmental differences that were observed earlier.

"School readiness" has been a concern of educators and policymakers for more than four decades, but little data have been available to assess that readiness across the national population of children. The Early Childhood Longitudinal Study of the Kindergarten Class of 1998-99 (ECLS-K) is addressing this need by following a nationally representative sample of children from kindergarten through fifth grade.

Below, performance by race/ethnicity, and by socioeconomic status (SES) of parents, is summarized for two aspects of reading proficiency and two aspects of mathematics proficiency. The data discussed here are drawn from the ECLS-K and based on an analysis from an earlier ETS Policy Information Report.[7]

Reading Proficiency

Recognizing the letters of the alphabet is one of a number of indicators of kindergarten reading proficiency. Among children entering kindergarten in the fall of 1998, a considerable proportion (65 percent) could recognize the letters of the alphabet — but there were differences by race/ethnicity, as shown in Figure 4.

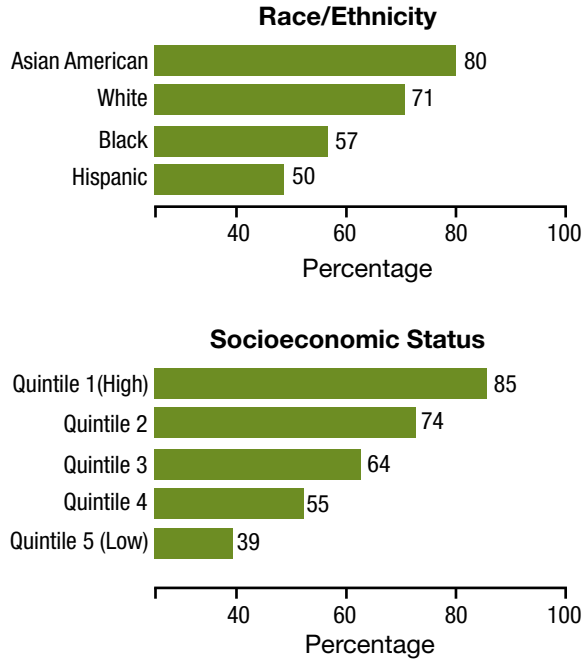
There were also considerable differences by the SES of the children's parents or guardians, ranging from 85 percent in the highest SES quintile, down to 39 percent in the lowest quintile. Lower-scoring minority children were also more likely to be in families with lower SES; when SES was held constant, almost all differences in reading readiness among racial/ethnic groups disappeared.

Also important in reading readiness is being able to understand the letter-sound relationship at the beginning of words. Among all kindergartners, 30 percent could understand such sounds; but again, there was considerable variation by race/ethnicity and by SES.

On the basis of race/ethnicity, the percentage ranged from 44 percent for Asian students, down to 20 percent for both Black students and Hispanic students, with White students coming in at 34 percent. Differences by SES ranged from 51 percent for the top quintile (highest SES), down to 10 percent for the bottom quintile (lowest SES).

Figure 4

Percentage of Kindergartners Who Could Recognize Letters of the Alphabet, by Race/ Ethnicity and Socioeconomic Status



Source: Richard J. Coley, *An Uneven Start: Indicators of Inequality in School Readiness*, Policy Information Report, Policy Information Center, Educational Testing Service, March 2002.

Three other aspects of reading proficiency were also assessed, all involving tasks more difficult for kindergartners than the two described above. For example:

- Just 17 percent could understand the ending sounds of words: 29 percent of Asian, 20 percent of White, 10 percent of Black, and 11 percent of Hispanic kindergartners.
- Only 2 percent could recognize common words: 9 percent of Asian, 3 percent of White, 1 percent of Black, and 1 percent of Hispanic kindergartners.
- Only 1 percent of kindergartners could understand words in context: 5 percent of Asian, 1 percent of White, and half a percent each of Black and Hispanic kindergartners.

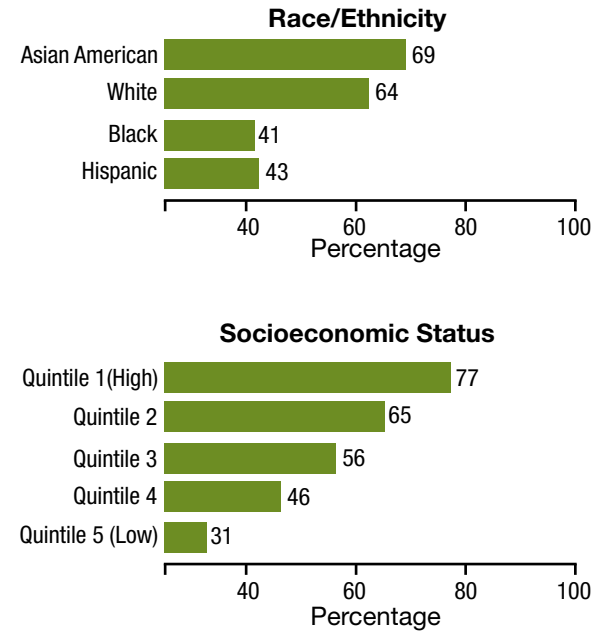
[7] Richard J. Coley, *An Uneven Start: Indicators of Inequality in School Readiness*, Policy Information Report, Policy Information Center, Educational Testing Service, March 2002. Socioeconomic status is measured from a scale that reflects the education, income, and occupations of kindergartners' parents or guardians. The scale is then divided into five quintiles for the purpose of making comparisons.

Mathematics Proficiency

Among the skills tested in mathematics was an understanding of the concept of relative size (e.g., reading all single-digit numbers, counting beyond 10, recognizing the sequence of patterns, and using non-standard units of length to compare objects). Fifty-seven percent of the kindergartners succeeded at tasks involving relative size, with considerable differences by race/ethnicity and socioeconomic status, as seen in Figure 5.

Figure 5

Percentage of Kindergartners Who Could Understand Relative Size, by Race/Ethnicity and Socioeconomic Status



Source: Richard J. Coley, *An Uneven Start: Indicators of Inequality in School Readiness*, Policy Information Report, Policy Information Center, Educational Testing Service, March 2002.

Among children whose parents or guardians were in the highest SES quintile, 77 percent showed an understanding of relative size; only 31 percent of kindergartners in the lowest SES quintile could do so.

Twenty-one percent of kindergartners succeeded in a task called "understanding ordinal sequence." This involved reading two-digit numerals, recognizing the next number in a sequence, identifying the ordinal position of an object, and solving a simple word problem. Among racial/ethnic groups, Asian students were the most likely to be able to do this task (32 percent).

In terms of SES, the percentages understanding ordinal numbers ranged from 39 percent at the highest quintile down to only 6 percent at the lowest.

On the easiest task assessed, recognizing numbers and shapes, almost all of the students succeeded, 93 percent over all. Minority students (Black and Hispanic) were only slightly lower. But when it came to manipulating these numbers, all of the students struggled.

- Just 4 percent could add and subtract: 9 percent of Asian, 5 percent of White, 1 percent of Black, and 2 percent of Hispanic kindergartners.
- Only 3 percent of all kindergartners could multiply and divide.

* *

For about 30 years we have known, from NAEP, that achievement varied by race/ethnicity and SES at the fourth-grade level, and that this unequal achievement continued through high school. We now know this inequality also exists at the kindergarten level — a fact that is not surprising, given the inequality we saw developing in the first three years of life.

In this era of test-based accountability, we have come to view school achievement through the window of cut points — specifically, the percentage of students reaching the *Proficient* level. Gaps in achievement are seen as differences in the percentage of subgroups who reach or exceed the standard cut point.

This approach may be effective in illuminating how much of the population has reached a designated standard, but in terms of measuring changes in achievement, it misses a lot. But the data reporting by NAEP enables the use of different lenses and indices to help explain performance.

For example, the percentage of students who score at the *Proficient* level on NAEP tells us only about the relatively few students who are just below and just above the cut point — the part of the score distribution where the change is reported from year to year. But this tells us nothing about students further down or further up the achievement distribution. In addition, it makes comparing gaps in achievement from one state to another difficult, because states often set the definition of proficiency at different points along the achievement scale.

When a low cut point is set, all subgroups may be deemed *Proficient*, resulting in no apparent achievement gap. A high cut point is more likely to result in a large achievement gap. In this way, it is possible for a state showing a small achievement gap to have lower average scores for minority students than a state showing a large achievement gap.

The use of average scores, while readily available from NAEP, has gone out of style somewhat in this era of standards. An average does have the merit of being derived from the scores of all students. Also, comparisons among subgroups can be made without concern for whether cut points have been set at different levels. This allows for easier comparisons of gaps.

While average scores provide easily understood indicators that summarize performance with a single number, we are reminded of the proverbial statistician who almost drowned in the lake that was, on average, only three feet deep. Averages can mask important differences within a population and provide no information about how that population performs at different points along the score distribution.

This section provides a number of different windows through which to view achievement and inequality at the national level, and the next section focuses on the state level. The data used are reported by NAEP, which has three ongoing programs: National and State NAEP (generally referred to as Main NAEP), the NAEP Trial Urban District Assessment, and Long-Term Trend. Main NAEP reaches back to 1990 and these assessments are based on subject matter frameworks developed by the National Assessment Governing Board, use the latest methodologies, and evolve as instructional practice changes.

For showing trends in average scores, we have used Long-Term Trend NAEP to provide the longest term perspective available — to the 1970s. These data report student performance at ages 9, 13, and 17. According to NAEP, "Measuring trends of student achievement or change over time requires the precise replication of past procedures and does not evolve based on changes in curricula or in educational practices, unlike the main NAEP national and state assessments."

We also used NAEP Long-Term Trend in the analysis of performance by quartiles, looking at changes in the achievement gap, since many dramatic changes in the achievement gap occurred before Main NAEP was started. The data in Figure 8 are from Main NAEP. The data in the other figures in this section are from Long-Term Trend NAEP.

We begin this section of the report with an examination of average scores on the reading and mathematics assessments of NAEP. This is followed by views of achievement based on achievement levels and followed by views of achievement based on quartiles. Organizing achievement by quartiles is consistent with national goals set at the education summit called by President George H.W. Bush nearly 20 years ago. These included that "the academic performance of all students at the elementary and secondary level will increase significantly in every quartile, and the distribution of minority students in each quartile will more closely reflect the student population as a whole."

Finally, scores at several different percentiles on the score scale are presented to provide a more panoramic view of student performance.

Using Average Scores

Average NAEP scores are common metrics, or windows, through which to view the achievement of U.S. students and the changes that have occurred over time. These scores provide a single indicator of student performance and can be useful in comparing the average performance of students from different racial/ethnic groups. Figures 6 and 7 show trends in the average reading and mathematics scores for the three age levels assessed by NAEP, breaking the data out by racial/ethnic group.[8]

[8] Data for Asian/Pacific Islander and American Indian/Alaskan Native students are not reported due to insufficient sample sizes.

Figure 6

Trends in Average NAEP Reading Scale Scores for Students Ages 9, 13, and 17

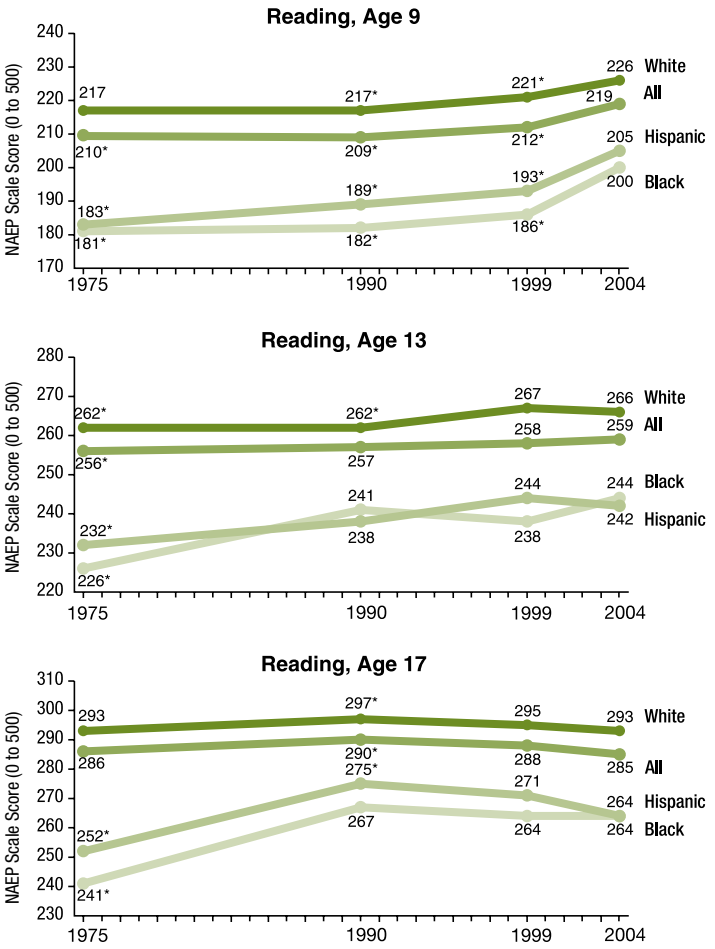
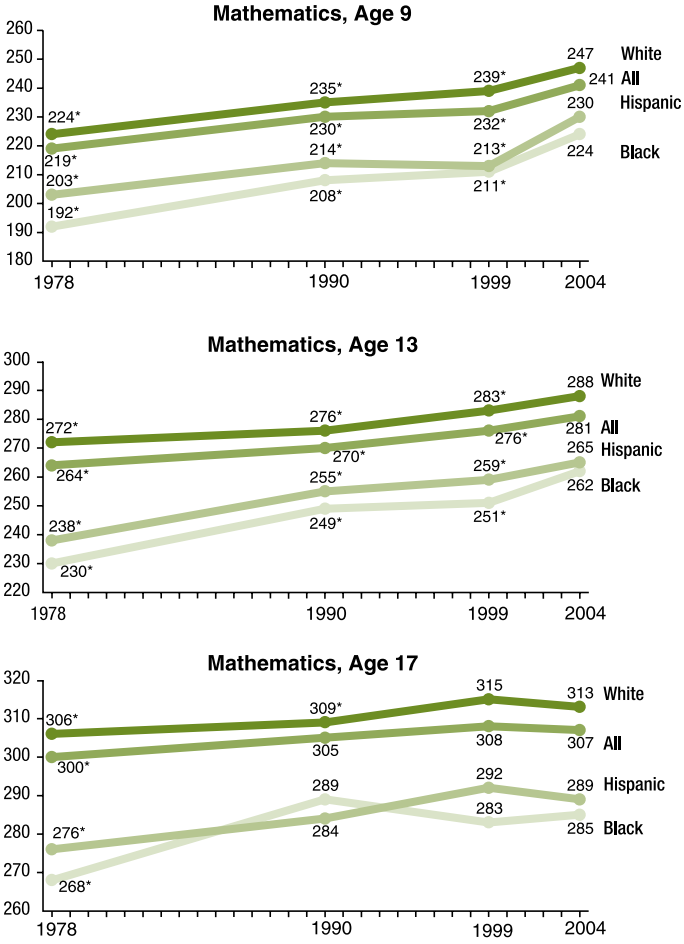


Figure 7

Trends in Average NAEP Mathematics Scale Scores for Students Ages 9, 13, and 17



* Significantly different from 2004
Source: Marianne Perie, Rebecca Moran, and Anthony D. Lutkus, *NAEP 2004 Trends in Academic Progress: Three Decades of Student Performance in Reading and Mathematics* (NCES 2005-464), U.S. Department of Education, IES, NCES, 2005.

Reading. In 2004, the nation's 9-year-olds attained the highest scores of any previous year. This improvement is seen for White, Black, and Hispanic students. Between 1975 and 2004, the average reading score for Black and Hispanic 9-year-olds increased by about 20 scale points, while the score for White students increased by about 10 points. For 13-year-olds, across racial/ethnic groups, scores in 2004 are up from 1975, but the only improvement made since 1990 is among White students. For all 17-year-olds, as well as for White and Hispanic subgroups, the 2004 reading score was lower than the 1990 score. However, substantial improvements were registered in the average scores of Black and Hispanic students from 1975 to 1990.

Mathematics. The average mathematics scores of 9- and 13-year-olds in all racial/ethnic groups were higher in 2004 than in any previous year (the difference in the score of Hispanic students between 1999 and 2004 was not statistically significant). For 17-year-olds, scores improved from 1978 but have been flat since 1990, with the exception of White students, who improved between 1990 and 2004.

Using Achievement Levels

NAEP results are reported at three different achievement levels: *Basic*, *Proficient*, and *Advanced*. These are performance standards reflecting where students are regarding what they should know and be able to do. In this section of the report, achievement-level results are presented as the percentage of students who score at or above *Proficient*, a level representing solid academic performance. Students who reach this level have demonstrated competency over challenging subject matter.[9]

Figure 8 shows trends in the percentage of students scoring at or above *Proficient* in reading and mathematics at grades 4 and 8 by racial/ethnic group.

Reading. In fourth-grade reading in 2007, 43 percent of White students scored at or above the *Proficient* level, compared with 17 percent of Hispanic students and 14 percent of Black students. While this gap is large, all three racial/ethnic groups examined here have shown statistically significant improvements on this measure over the past 15 years. Among eighth-graders, 40 percent of White students scored at or above *Proficient*, compared with 15 percent of Hispanic and 13 percent of Black eighth-graders. The trend lines for all three groups are flat; there has been no improvement on this measure for White, Black, or Hispanic eighth-graders in the past 15 years.

Mathematics. In mathematics, there is both good news and bad news. For all three racial/ethnic groups at both grade levels, the percentage of students scoring at or above *Proficient* increased at each assessment interval between 1990 and 2007. The increases between 1990 and 2007 are particularly striking for fourth-graders. For example, in 1990, just 1 percent of Black fourth-graders scored at or above *Proficient*. By 2007 that percentage had increased to 15. For Hispanic students the percentage increased from 5 to 22, and for White students the percentage increased from 16 to 51. The bad news is that there is a very large gap between White and minority students and that gap is increasing. Among eighth-graders, for example, 42 percent of White students scored at or above *Proficient* in 2007, compared with 15 and 11 percent, respectively, of Hispanic and Black eighth-graders.

Viewing the Distribution of Scores

While average scores and achievement levels provide two windows into understanding achievement and inequality, a more panoramic view is provided by examining how achievement is distributed over the entire score range. This can be accomplished by arraying NAEP scores at various percentiles.

Percentiles show the percentage of students whose scores fall below a specified point on the NAEP scale. So the score at the 90th percentile is the score at which 90 percent of students score below — or, conversely, the score that the top 10 percent of students score above. Viewing percentiles can help us see whether changes in average scores, such as those shown in Figures 6 and 7, are reflected at different parts of the score distribution.

For example, we can determine how the top scorers have fared over time or whether the lowest-scoring students have improved. This is an illuminating window. It can help us understand whether rising average scores are the result of improvement by only the highest- or lowest-scoring students, or whether average improvement is the result of increases all along the score distribution.

Figure 9 shows percentile results for reading and math at three age levels, for 1990 and 2004. We also show the data separately for White, Black, and Hispanic students. In reading, at age 9, there were score increases for the total group of students at the 50th, 25th, and 10th percentiles — showing that increases during that period came mainly from the middle and bottom parts of the score distribution. Black 9-year-olds, however, registered increases all along the score distribution; and Hispanic 9-year-olds showed improvements at all but the 90th percentile.

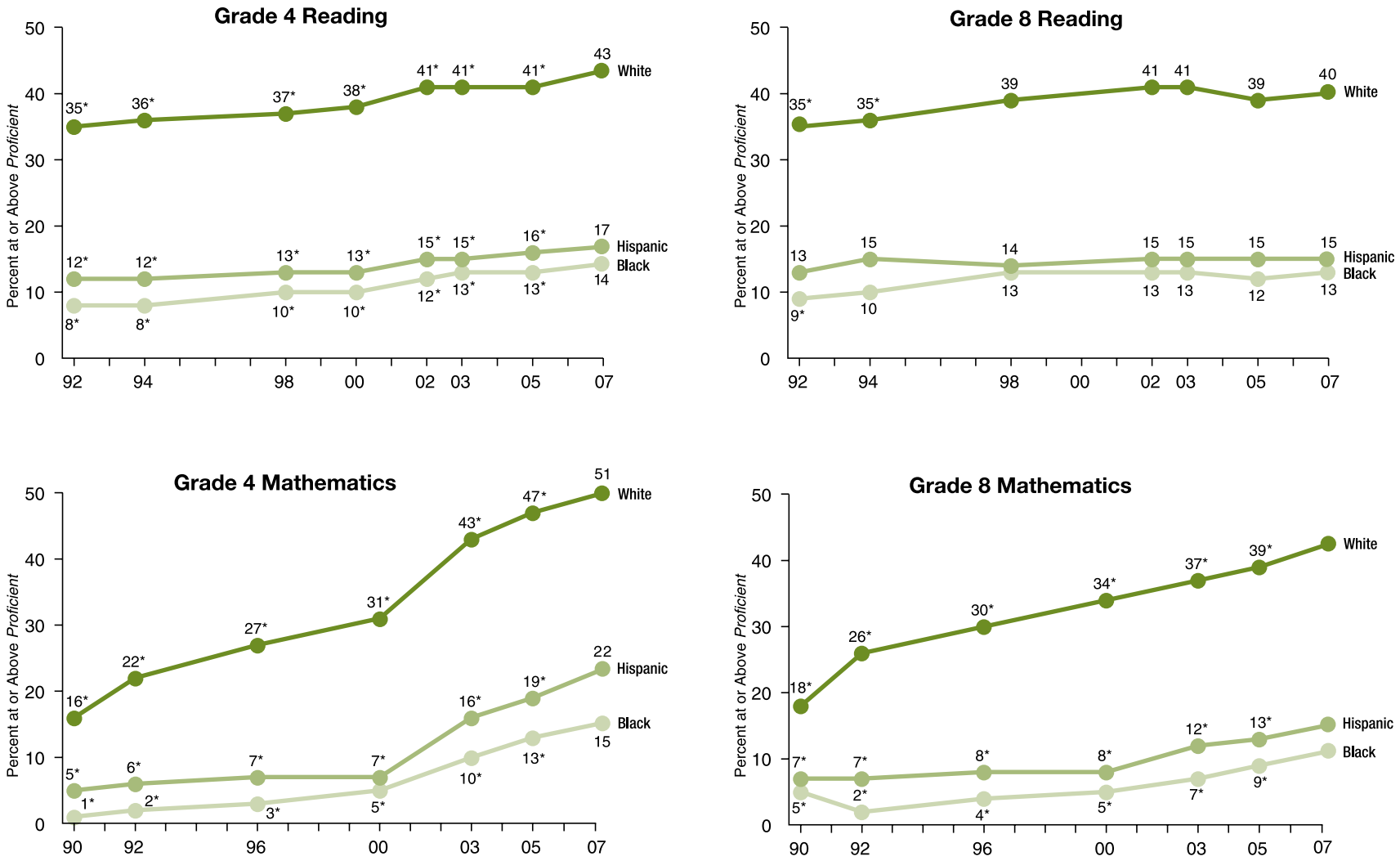
The good news in reading scores does not carry over to older students, however. There was little change for 13-year-olds over the time period shown. And among 17-year-olds, there were drops at both ends of the distribution scale — at the 75th, 25th, and 10th percentiles.

The news is better in mathematics, at least for 9- and 13-year-olds. White, Black, and Hispanic 9-year-olds registered gains at nearly every percentile level. For 17-year-olds, some improvement can be seen at the bottom of the score distribution for the total population and for White students. There were no statistically significant changes for Black or Hispanic students.

[9] Achievement levels are set by the National Assessment Governing Board, based on recommendations from panels of educators and members of the public, to provide a context for interpreting student performance on NAEP. Detailed descriptions of the NAEP achievement levels can be found at http://www.nagb.org/pubs/pubs.html.

Figure 8

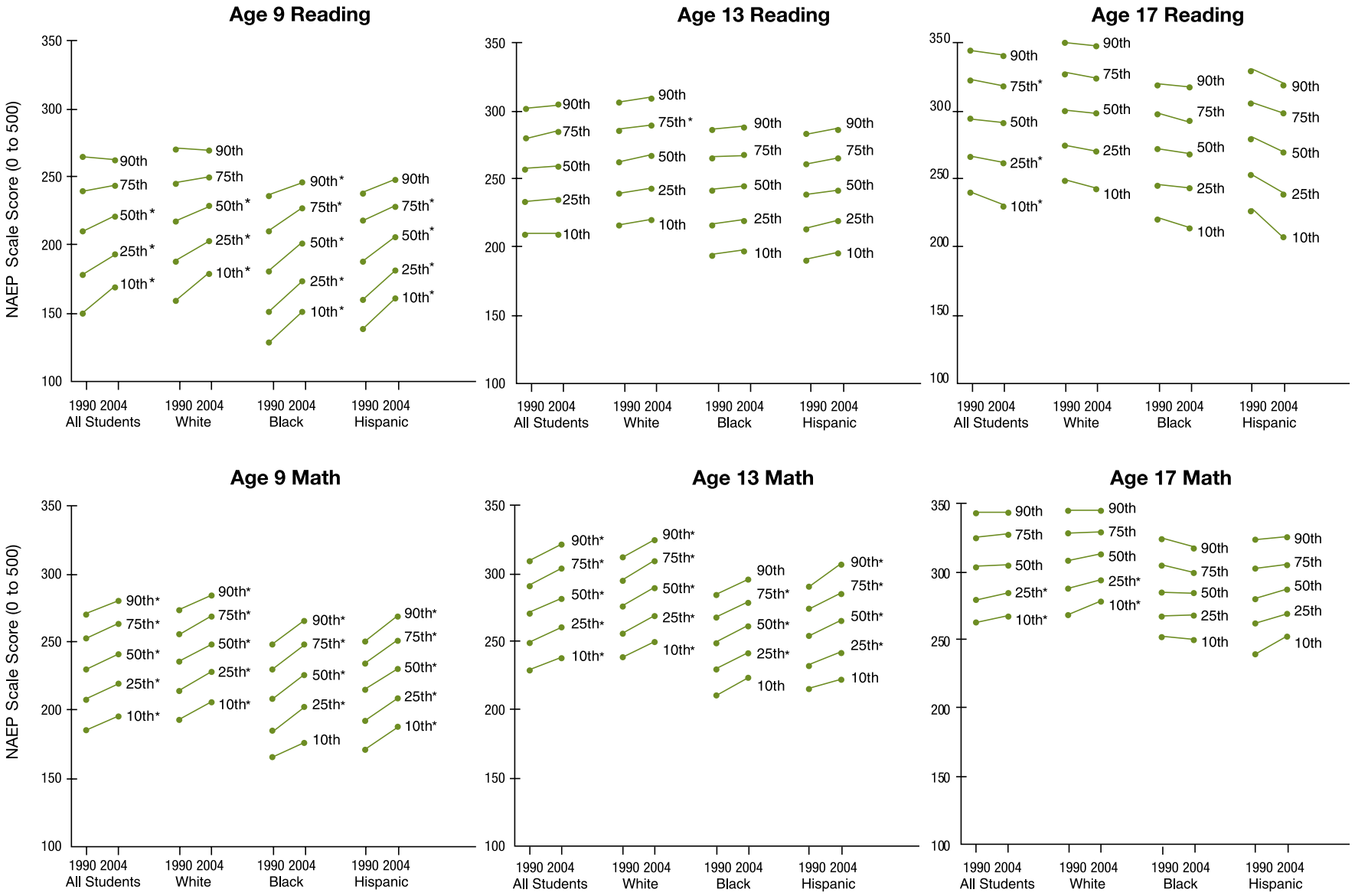
Trends in the Percentage of Students at or Above* Proficient *in NAEP Reading and Mathematics, Grades 4 and 8, by Racial/Ethnic Group



Note: Accommodations not permitted in 1992 and 1994.
*Significantly different from 2007
Source: Data from the National Assessment of Educational Progress analyzed by the ETS Policy Information Center.

Figure 9

Percentile Distribution of NAEP Reading and Mathematics Scores, by Age and Racial/Ethnic Group, 1990 and 2004



* Indicates statistically significant difference from 1990 to 2004
Source: NAEP special tabulations prepared by ETS.

Examining Quartiles

Another way to look at test-score data (and the final one that we will discuss here) is to examine changes in score distributions by breaking them into quartiles. We can see, for example, how the scores of students in quartile 4 (the highest) in one year compare with those of students in quartile 4 in previous years. This method also allows us to see how the lowest-scoring group of students is performing over time. Such information can help determine whether changes in average scores are occurring because of improvements at lower score levels or because of improvements by the highest-scoring students. A view through this window, for example, might show us that all the improvement in mathematics scores came at the bottom of the score distribution, offering the possibility that a focus on basic skills might be responsible for the gains.

Reading. Figures 10, 11, and 12 show a quartile analysis of NAEP reading scores over time for all students, White students, and minority students.[10] The bottom section of the charts shows data on the achievement gap broken out by quartile.

For 9-year-olds, there were score increases at each quartile, indicating relatively even growth. Adding to this good news is the decline in the achievement gap that can be seen in the bottom of the chart. In contrast to the picture seen using Main NAEP results, in the Long-Term Trend data each quartile shows progress in closing the gap between White and minority students. On the other hand, we still see a substantial score gap of between 21 and 27 scale points across quartiles.

For 13- and 17-year-olds, the picture is much more static. At age 13, the small score changes seen at each quartile between 1999 and 2004 are not statistically significant, except for the increase in the third quartile for minority students. We also see some closing of the gap among 13-year-olds between 1975 and 1990 at all quartiles. Since 1990, however, little has been happening.

At age 17, the big change was among minority students, whose scores jumped at each quartile between 1975 and 1990; their scores have not improved since, however, and fell significantly at the lowest quartile in 2004. Similarly, there has been some progress in closing the gap among 17-year-olds — but it has all occurred between 1975 and 1990 and has been relatively level since.

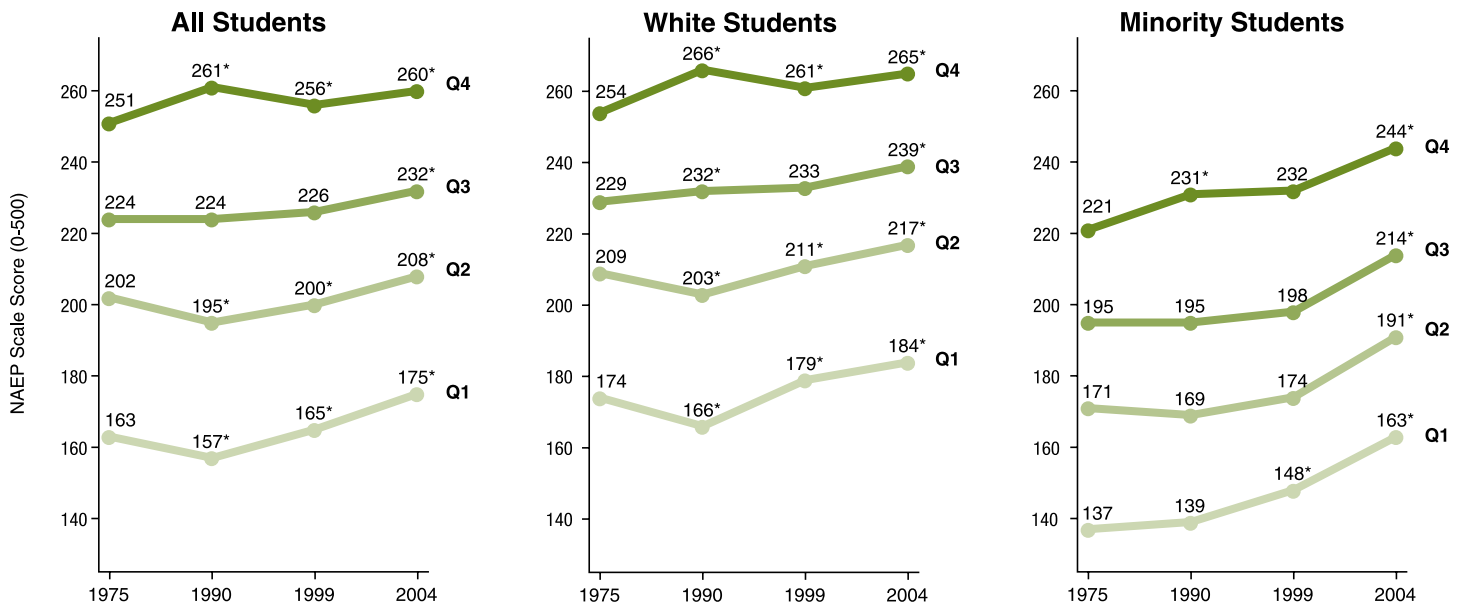
Mathematics. Figures 13, 14, and 15 provide a quartile analysis of NAEP mathematics scores over time for all students, White students, and minority students. For 9-year-olds, there has been substantial improvement at all quartiles since 1978. And for minority students, in particular, improvements between 1999 and 2004 are substantial at each quartile. Additionally, the score gap between White and minority 9-year-olds has narrowed at each quartile between 1999 and 2004.

There is also improvement shown among 13-year-olds at nearly all quartiles. Further, the achievement gap narrowed between 1978 and 1990 at all four quartiles and has continued to close at the highest two quartiles. For 17-year-olds, there has been little improvement at any quartile, with the exception of minority students. For these students, there was substantial improvement at all quartiles between 1978 and 1990; since then, however, the lines have remained basically flat. In terms of the gap, the most narrowing period came between 1978 and 1990. Since then, the gap has actually increased in several quartiles.

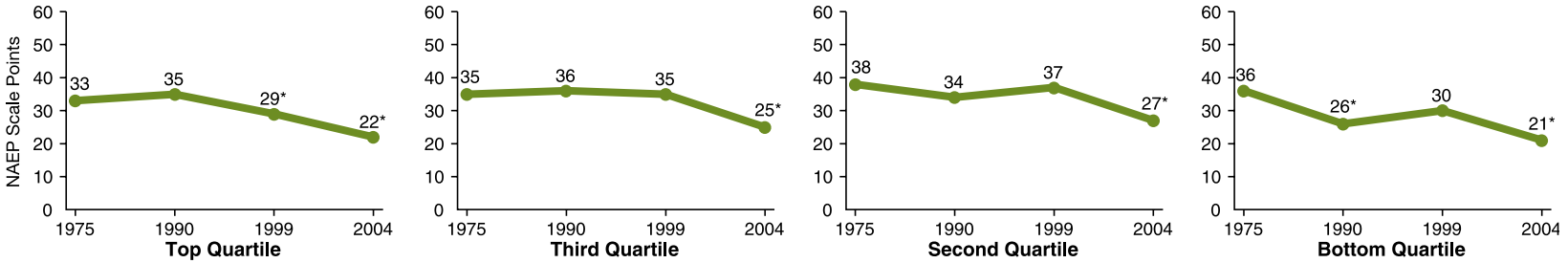
[10] In these charts, "minority students" are defined as Black and Hispanic students combined. While less than ideal, combining these two groups produces a larger sample size, which is important when dividing each group into fourths.

Figure 10

Trends in NAEP Reading Scores, by Quartile, Age 9



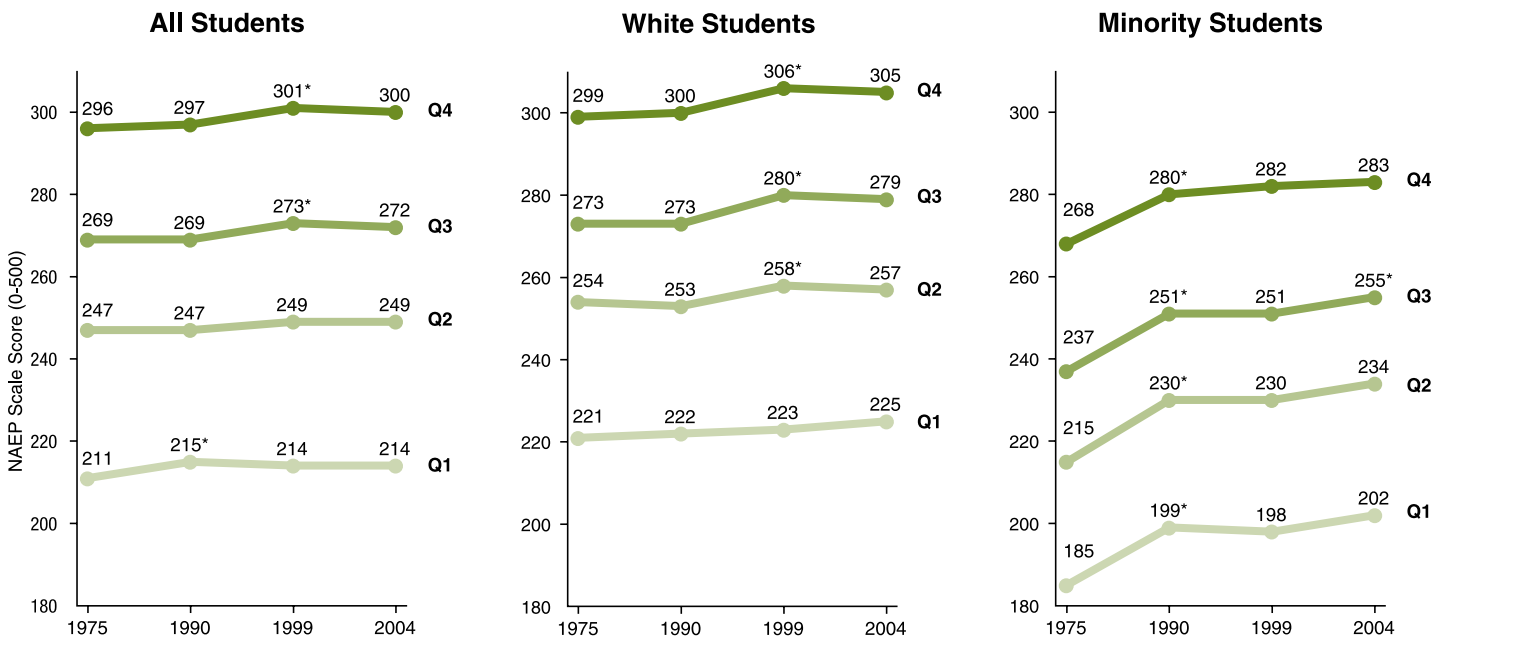
Trends in the NAEP Reading Test-Score Gap Between White and Minority Students, by Quartile, Age 9



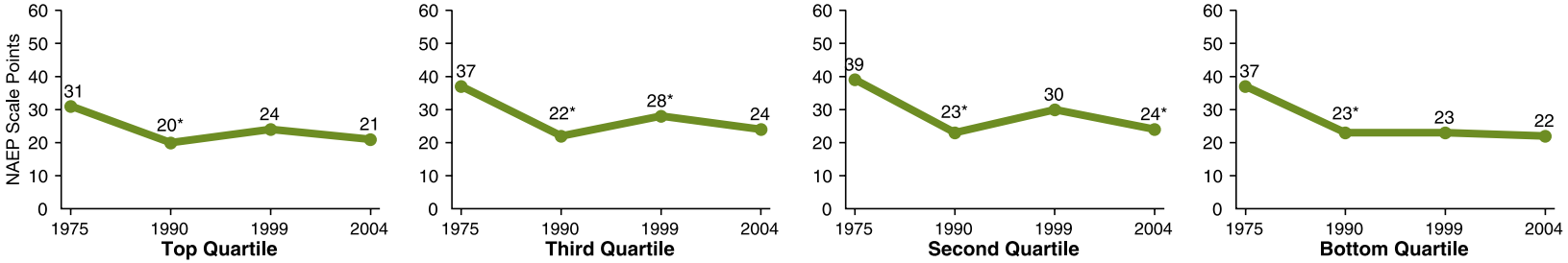
*Significant difference from previous comparison year
Source: NAEP Long-Term Trend data analyzed by ETS.

Figure 11

Trends in Average NAEP Reading Scores, by Quartile, Age 13



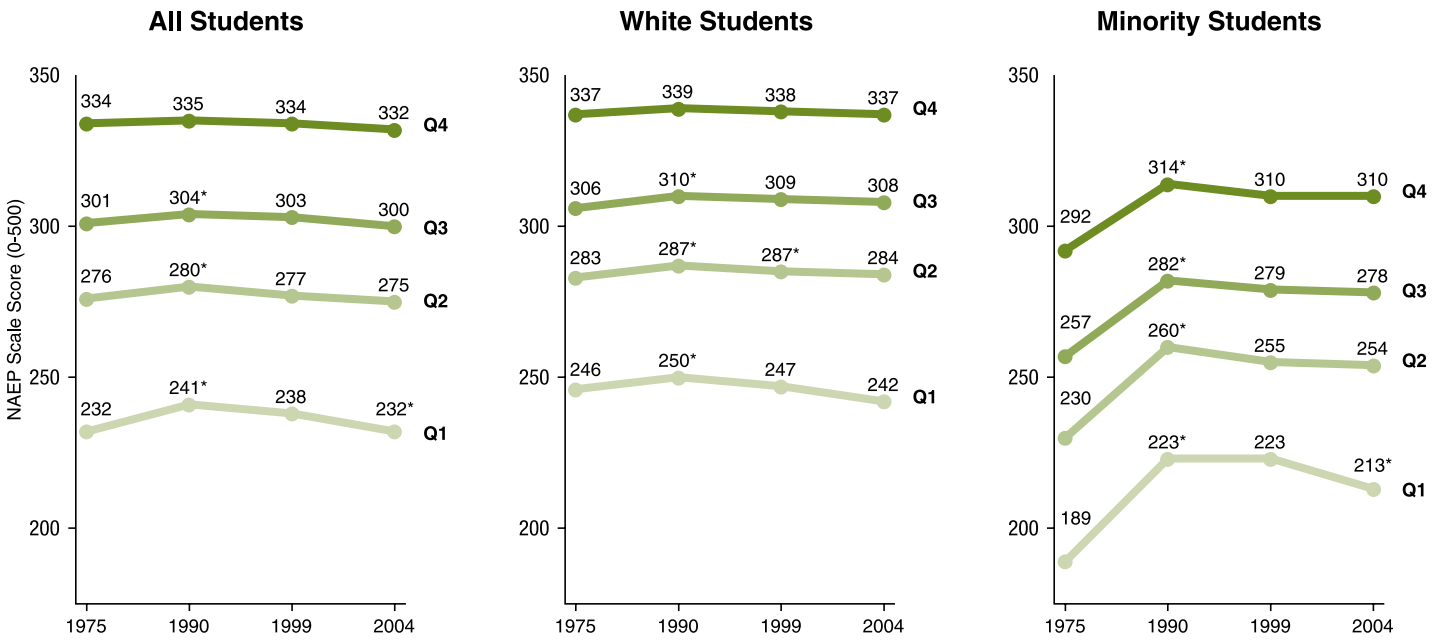
Trends in the NAEP Reading Test-Score Gap Between White and Minority Students, by Quartile, Age 13



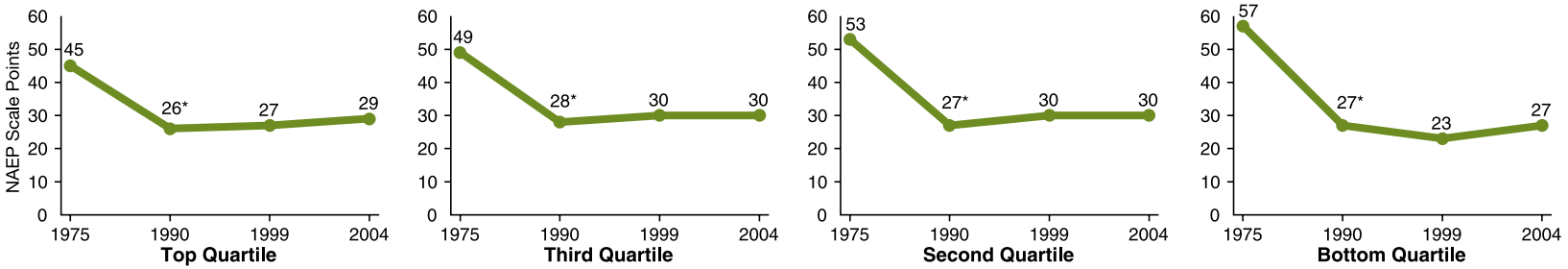
*Significant difference from previous comparison year
Source: NAEP Long-Term Trend data analyzed by ETS.

Figure 12

Trends in Average NAEP Reading Scores, by Quartile, Age 17



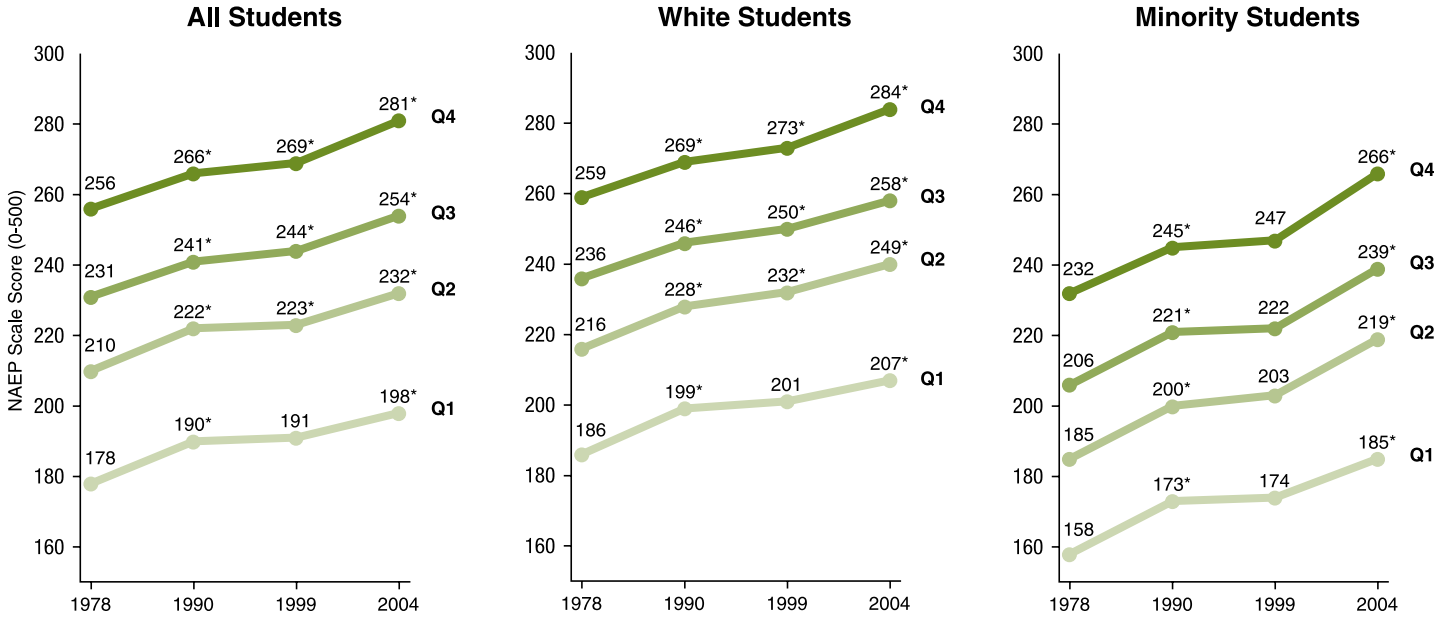
Trends in the NAEP Reading Test-Score Gap Between White and Minority Students, by Quartile, Age 17



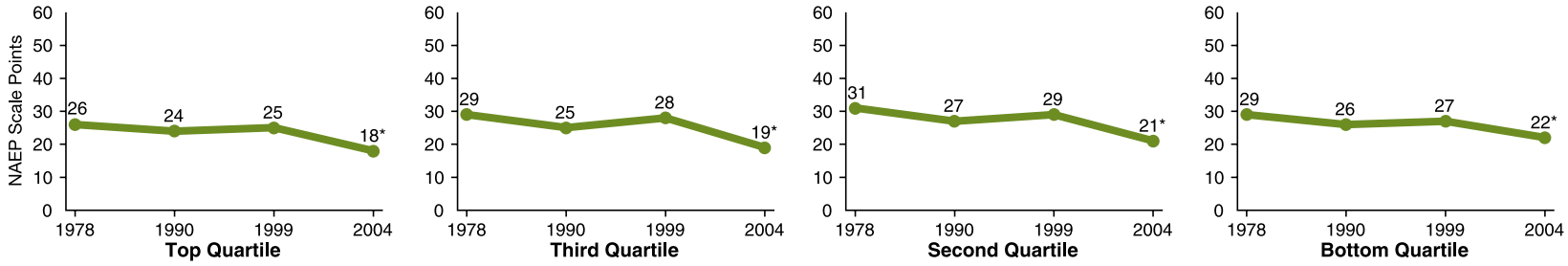
*Significant difference from previous comparison year
Source: NAEP Long-Term Trend data analyzed by ETS.

Figure 13

Trends in Average NAEP Mathematics Scores, by Quartile, Age 9



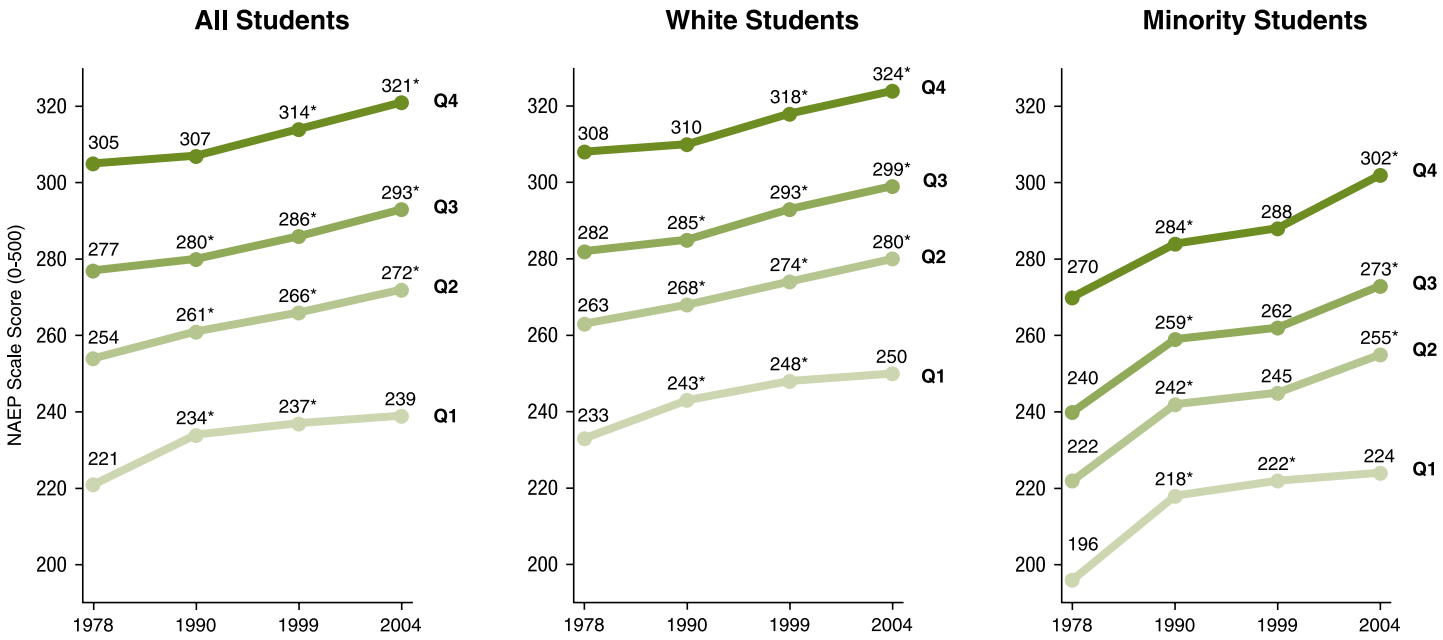
Trends in the NAEP Mathematics Test-Score Gap Between White and Minority Students, by Quartile, Age 9



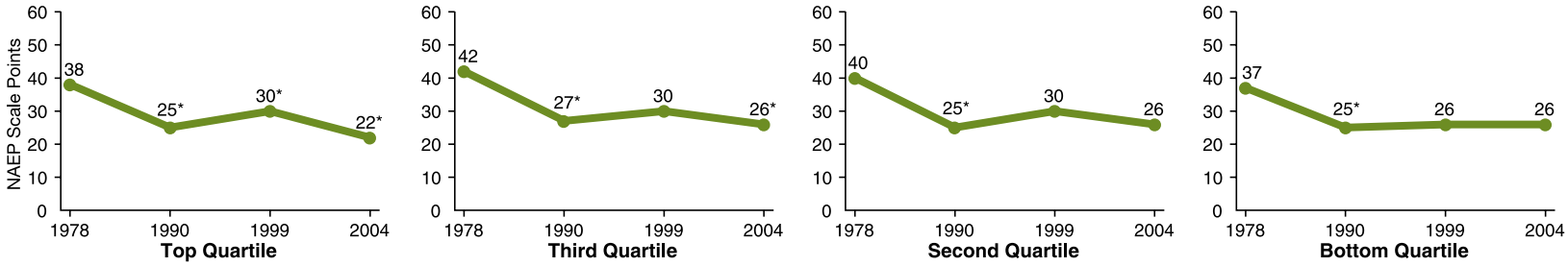
*Significant difference from previous comparison year
Source: NAEP Long-Term Trend data analyzed by ETS.

Figure 14

Trends in Average NAEP Mathematics Scores, by Quartile, Age 13



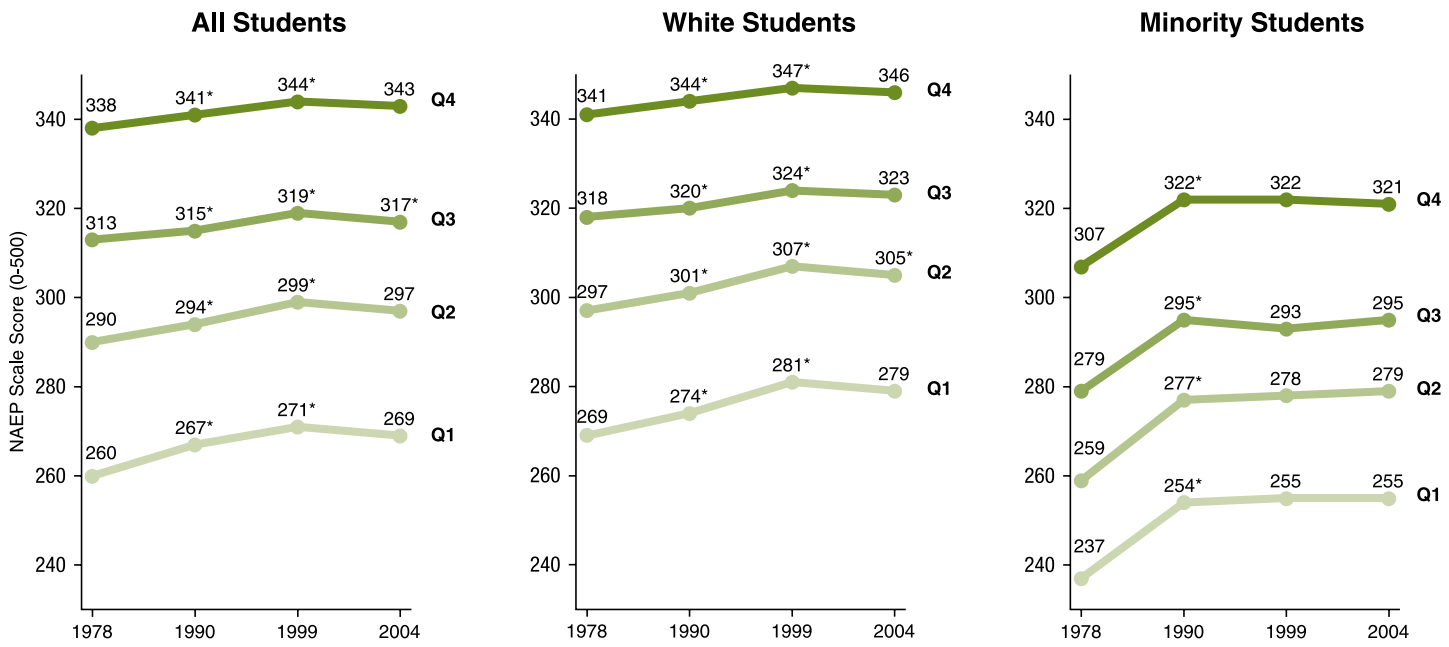
Trends in the NAEP Mathematics Test-Score Gap Between White and Minority Students, by Quartile, Age 13



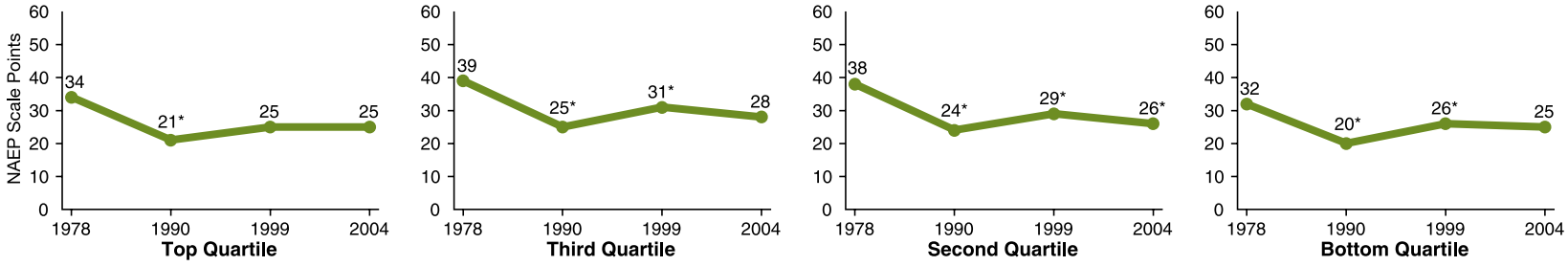
*Significant difference from previous comparison year
Source: NAEP Long-Term Trend data analyzed by ETS.

Figure 15

Trends in Average NAEP Mathematics Scores, by Quartile, Age 17



Trends in the NAEP Mathematics Test-Score Gap Between White and Minority Students, by Quartile, Age 17



*Significant difference from previous comparison year
Source: NAEP Long-Term Trend data analyzed by ETS.

The national estimates of achievement provided in the preceding section of this report are based on measures that have been aggregated across the total sample of students from the nation's schools and states. Since 2003, states have been required to participate in NAEP to receive Title 1 funding. Because of this, we can disaggregate the national data somewhat and see how individual states perform in reading and mathematics. And since many states participated in NAEP before 2003, we can examine changes over time.[11]

This section of the report uses data from Main NAEP to examine student achievement across states through the same lenses used to view national results — average scores, the percentage of students scoring at the *Proficient* level, and performance at the top and bottom of the score distribution.[12] It is important to keep in mind that these measures do not take into account differences among states' demographic profiles, nor do they consider differences in state policies regarding which students are or are not excluded from the assessments.

The advantages and limitations of focusing on the percentage of students who reach the *Proficient* cut point were discussed earlier in this report. The analyses presented here will show contrasts among different measures in how well states are performing.

Reading Results

Table 1 provides an overall summary of how states performed on the NAEP eighth-grade reading assessment between 2002 and 2007 and Table 2 lists the states that showed improvements. There was little good news: no states showed improvement in either the average score or the percentage of students scoring at the *Proficient* level. Five states improved in the top quartile score and five improved in the bottom quartile score. On the other hand, 14 states declined in the bottom quartile score, 12 declined in average score, 11 declined in the top quartile score, and 3 declined in the percentage of students reaching the *Proficient* level. Most states, however, showed no significant change between 2002 and 2007 on the four measures.

District of Columbia, Massachusetts, Nevada, New York, and Pennsylvania showed gains in the top quartile score, while Georgia, Maine, Maryland, North Dakota, and Pennsylvania showed gains in the bottom quartile score. Table 3 lists the states that showed declines between 2002 and 2007 on each of

Table 1

Overall Summary of States' Status on Four Measures, NAEP Reading, Grade 8

Grade 8 Reading Change, 2002 to 2007 in:	States Improving	States Unchanged	States Worse
Average Score	0	31	12
Percent *Proficient*	0	40	3
Top quartile score*	5	27	11
Bottom quartile score*	5	24	14

Note: Includes those states participating in both assessments, the District of Columbia, and DoDEA schools
*Quartile score is calculated as the average score of students in the quartile.

Table 2

States Showing Improvement Between 2002 and 2007 in NAEP Reading, Grade 8

Gain in Average Score	Gain in Percent *Proficient*	Gain in Top Quartile	Gain in Bottom Quartile
None	None	District of Columbia	Georgia
		Massachusetts	Maine
		Nevada	Maryland
		New York	North Dakota
		Pennsylvania	Pennsylvania

Table 3

States Showing Declines Between 2002 and 2007 in NAEP Reading, Grade 8

Decline in Average Score	Decline in Percent *Proficient*	Decline in Top Quartile	Decline in Bottom Quartile
Delaware	Kentucky	Idaho	Arizona
Kentucky	Michigan	Kansas	Delaware
Michigan	West Virginia	Michigan	Kentucky
Mississippi		New Mexico	Michigan
Missouri		North Carolina	Mississippi
Nebraska		North Dakota	Missouri
New Mexico		Oregon	Nebraska
North Carolina		Tennessee	New Mexico
Oklahoma		Utah	New York
Rhode Island		Washington	North Carolina
Washington		West Virginia	Oklahoma
West Virginia			Oregon
			Rhode Island
			West Virginia

Appendix Table 1 shows the reading results for all participating states.

[11] We also include data for the District of Columbia and Department of Defense Education Activity (DoDEA) overseas and domestic schools.
[12] All of the data in this section are based on analyses of data from the National Assessment of Educational Progress conducted by Educational Testing Service. All differences noted are statistically significant.

Table 4

Overall Summary of Changes on Four Measures, NAEP Mathematics, Grade 8

Grade 8 Mathematics Change, 2000 to 2007 in:	States Improving	States Unchanged	States Worse
Average Score	36	5	0
Percent *Proficient*	32	9	0
Top quartile score*	39	2	0
Bottom quartile score*	34	7	0

Note: Includes those states participating in both assessments, the District of Columbia, and DoDEA schools
*Quartile score is calculated as the average score of students in the quartile.

the four measures. Twelve states declined in average score, three declined in the percentage of students scoring at the *Proficient* level, 11 states declined in the top quartile score, and 14 declined in the bottom quartile score.

Mathematics Results

There was much more good news in mathematics than in reading. Tables 4 and 5 summarize the results seen between 2000 and 2007. The majority of participating states improved on all four measures examined. Thirty-six states improved in average score, 32 in percent *Proficient*, 39 in top quartile score, and 34 in bottom quartile score. Far fewer states showed no change over the period, and none of the participating states declined on any of the measures. Appendix Table 2 shows the mathematics results for all four measures for the participating states.

These data paint quite different pictures for states in reading and mathematics achievement. In reading, there was little good news between 2002 and 2007 — only a few states showed any improvement, a larger number of states showed declines, and most states showed no significant change for any of the measures examined. In mathematics, on the other hand, improvements were widespread for all of the measures examined. In addition, no states showed declines between 2000 and 2007. These data may be helpful to state policymakers and educators in taking note of where among these four measures they gained or lost ground.

Table 5

States Showing Improvement Between 2000 and 2007 in NAEP Mathematics, Grade 8

Gain in Average Score	Gain in Percent *Proficient*	Gain in Top Quartile	Gain in Bottom Quartile
Arizona	Arizona	Arizona	Alabama
Arkansas	Arkansas	Arkansas	Arizona
California	California	California	Arkansas
District of Columbia	District of Columbia	Connecticut	California
DoDEA	DoDEA	District of Columbia	District of Columbia
Georgia	Georgia	DoDEA	DoDEA
Hawaii	Hawaii	Georgia	Georgia
Idaho	Idaho	Hawaii	Hawaii
Illinois	Illinois	Idaho	Idaho
Indiana	Indiana	Illinois	Illinois
Kansas	Kansas	Indiana	Kansas
Kentucky	Kentucky	Kansas	Kentucky
Louisiana	Louisiana	Kentucky	Louisiana
Maine	Maryland	Louisiana	Maine
Maryland	Massachusetts	Maine	Maryland
Massachusetts	Mississippi	Maryland	Massachusetts
Minnesota	Missouri	Massachusetts	Minnesota
Mississippi	Nebraska	Michigan	Mississippi
Missouri	Nevada	Minnesota	Missouri
Nebraska	New Mexico	Mississippi	Nevada
Nevada	New York	Missouri	New Mexico
New Mexico	North Carolina	Montana	New York
New York	North Dakota	Nebraska	North Carolina
North Carolina	Ohio	Nevada	North Dakota
North Dakota	Rhode Island	New Mexico	Oklahoma
Ohio	South Carolina	New York	Rhode Island
Oklahoma	Tennessee	North Carolina	South Carolina
Rhode Island	Texas	North Dakota	Tennessee
South Carolina	Utah	Ohio	Texas
Tennessee	Vermont	Oklahoma	Utah
Texas	Virginia	Oregon	Vermont
Utah	Wyoming	Rhode Island	Virginia
Vermont		South Carolina	West Virginia
Virginia		Tennessee	Wyoming
West Virginia		Texas	
Wyoming		Utah	
		Vermont	
		Virginia	
		Wyoming	

Appendix Table 2 shows the mathematics results for all of the participating states.

Differences in the Achievement Gap Between White and Black Students Among the States

The previous state comparisons show different ways to look at student achievement but do not address achievement gaps by race/ethnicity. References to the size of gaps among states are typically put in terms of the percent reaching or exceeding the *Proficient* cut point. Table 6 illustrates a more complete review of such gaps, using the NAEP assessment of eighth-grade mathematics for 2007. Achievement gaps vary, depending on whether the measure is the average scale score, or the percentage reaching the *Proficient* achievement level, or the percentage reaching the *Basic* achievement level. Each of these measures is shown in the table, along with the state's rank on the particular measure.

An advantage of using the difference in the average scale score is that this measure includes the scores of all students in its calculation. The first column in Table 6 shows the difference between the average score of White and Black students and ranks the states on this number from low gap to high gap. On this measure, Oregon ranks at the top of the list with a White-Black gap of 16 points, while Nebraska ranks at the bottom with a gap of 51 points between White and Black students.

The size of the gap in the percentage of students who reach particular cut points, or achievement levels, on the NAEP scale will vary depending on the cut point or achievement level chosen for comparison. The second set of columns in Table 6 shows the gap in the percentage of White and Black students who score at or above the *Proficient* level. Oregon, again, remains at the top of the table with the smallest gap, but the rankings of some other states change considerably. Department of Defense Education Activity overseas and domestic schools (DoDEA), for example, change from a ranking of 2 on average score gap, to a ranking of 10 on the gap in the percentage of students scoring at or above *Proficient*. Massachusetts drops down to the bottom rank on this measure, reporting a gap of 45 percentage points between White and Black students. There is considerable movement of many states in these rankings.

Overall, the gap in the percentage of White and Black eighth-graders who score at or above the *Basic* level is higher than the gap for the other two measures. The average gap for the states included here is 35 points. Oregon continues to rank highest among the states, reporting a gap of only 19 points, and DoDEA schools climb back up to second with a gap of 21 points. Massachusetts is replaced by Nebraska at the bottom of the rankings on this measure. Again, there is considerable movement among many of the participating states on this measure.

Although the picture changes relatively little at the top and bottom of these rankings, there is considerable shifting that occurs throughout the table. In addition to the examples provided above, West Virginia, Oklahoma, and Mississippi rank higher on average score and percent of students scoring at or above *Proficient* than they do on the percentage of students scoring at or above *Basic*. Texas, on the other hand, ranks considerably higher on the percentage of its student scoring at or above the *Basic* level than it does on the other two measures.

These data demonstrate that the size of achievement gaps can be pinned down, states can be compared, and changes in gaps can be tracked. But it is always necessary to know what measure is being used to determine the size of a gap.

Sometimes we see jurisdictions with small gaps, but the reason is that scores are low. And when we look at changes in gaps over time, we see "good" gap closings (in which all groups are performing better, but minorities are improving at a greater rate) and "bad" gap closings (in which White and minority students' scores are both going down, but one at a greater rate).

Table 6

State Rankings on the Achievement Gap Between White and Black Eighth-Graders Using Three NAEP Mathematics Measures, 2007

	Average Score		At or Above *Proficient*		At or Above *Basic*	
State	Rank	Difference Between White and Black Students (in Scale Points)	Rank	Difference Between White and Black Students (in Percentage Points)	Rank	Difference Between White and Black Students (in Percentage Points)
Oregon	1	16	1	11	1	19
DoDEA	2	19	10	26	2	21
New Mexico	3	21	6	21	4	25
West Virginia	3	21	2	15	10	32
Oklahoma	4	22	3	16	9	31
Arizona	5	23	9	25	3	23
Kentucky	6	25	4	18	8	30
Louisiana	6	25	6	21	11	35
Nevada	7	26	5	20	7	29
Georgia	8	27	10	26	9	31
South Carolina	8	27	11	29	6	28
Arkansas	9	28	7	22	9	31
Kansas	9	28	12	30	8	30
Mississippi	9	28	5	20	15	39
Tennessee	9	28	8	23	12	36
Virginia	9	28	14	32	8	30
North Carolina	10	29	14	32	9	31
Texas	10	29	19	37	5	26
United States		31		30		35
Alabama	11	32	8	23	15	39
Indiana	11	32	14	32	11	35
New York	11	32	11	29	12	36
Ohio	12	33	15	33	12	36
Missouri	13	34	13	31	18	43
Rhode Island	13	34	10	26	13	37
California	14	35	11	29	16	40
Maryland	15	36	20	40	11	35
Minnesota	16	37	17	35	14	38
Connecticut	17	38	18	36	15	39
Illinois	17	38	16	34	17	41
Massachusetts	18	40	21	45	13	37
Michigan	19	41	13	31	19	48
Nebraska	20	51	18	36	20	54

Excludes states with samples too small to support estimates of the gap.

The history of testing in the United States is one of measuring students at one point in time and then comparing them with previous students in the same grade. There is now great interest in measuring growth in knowledge of the same students over the course of the school year — how much they grew, in addition to what they know, from experiences both in and outside of school.

Official reports of NAEP tell us what students over the years know in the winter of fourth, eighth, and 12th grades (or at ages 9, 13, and 17). While NAEP does not report how much students grow in knowledge from the fourth to the eighth grade, for example, the fact that the assessments have been given four years apart and the scores reported on a common scale enables us to use the data to estimate how much students' knowledge has grown over the four years.[13] Although this assessment pattern has not always been followed precisely, it has been used closely enough to permit some comparisons based on following a cohort of students during a four-year period of schooling.[14]

Reading

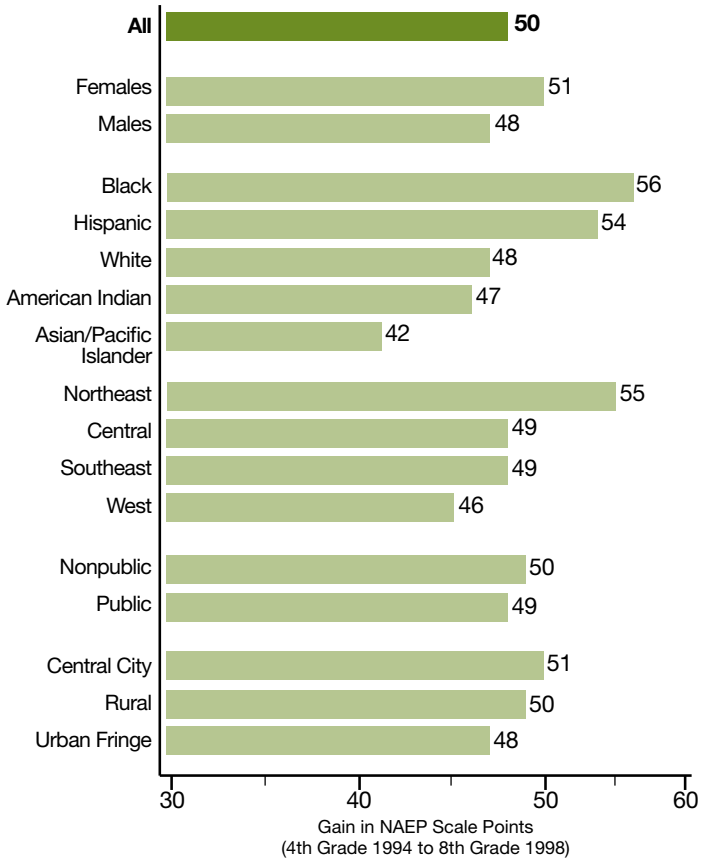
There is a stark contrast between the kind of subgroup differences seen in traditional NAEP reporting of what, in total, students know and how much their knowledge grew over the four-year period from fourth to eighth grade. Figure 16 shows the cohort growth in learning for reading, from grade four in 1994 to grade eight in 1998.[15]

Over this period, the average growth for all students was 50 points on the NAEP achievement scale of 0 to 500 — 51 points for females and 48 points for males. Black students grew the most (56 points), while Asian/Pacific Islander students grew the least (42 points). White students were third at 48 points, just below the average for all students.[16]

These data are in stark contrast with the large differences in average scores representing what students knew at the end of fourth and eighth grades. On this measure, Asian/Pacific Islander students score the highest, followed by White students, with Black and Hispanic students trailing well below. When students start school, there is wide variation in what they know and can do, and in their cognitive development generally.

Figure 16

Cohort Growth in NAEP Reading, Grades 4 to 8, by Standard NAEP Reporting Groups



Source: Data from the National Assessment of Educational Progress analyzed by the ETS Policy Information Center.

If they grow at somewhat similar rates, these large differences will prevail throughout school.

There is also a striking contrast among states in terms of gains in average NAEP scores between fourth- and eighth-graders — and there are large differences in how states are ranked, depending on which measure is used. For example, in grade eight, NAEP scores showed Maine on top in "level of knowledge," with an average score of 273 on the 0 to 500 scale; however, it placed fourth from the bottom in terms of gain in achievement.

[13] Although it is unlikely that many of the same students would be assessed at both grade levels, students assessed constitute a representative sample of approximately the same group of students.

[14] For additional information on the statistical and measurement challenges posed in following a cohort of students using NAEP data, see Richard J. Coley, *Growth in School Revisited: Achievement Gains from the Fourth to the Eighth Grade*, Policy Information Report, Policy Information Center, Educational Testing Service, 2003. These challenges include, for example, the possibility of changes in exclusion rates (states can exclude students from the assessments because of language difficulties or handicapping conditions) or changes in cohort composition because of student mobility in or out of a particular state.

[15] The data in this section of the report are drawn from Coley, 2003.

[16] The difference between Black, White, and Asian/Pacific Islander students is statistically significant.

Mathematics

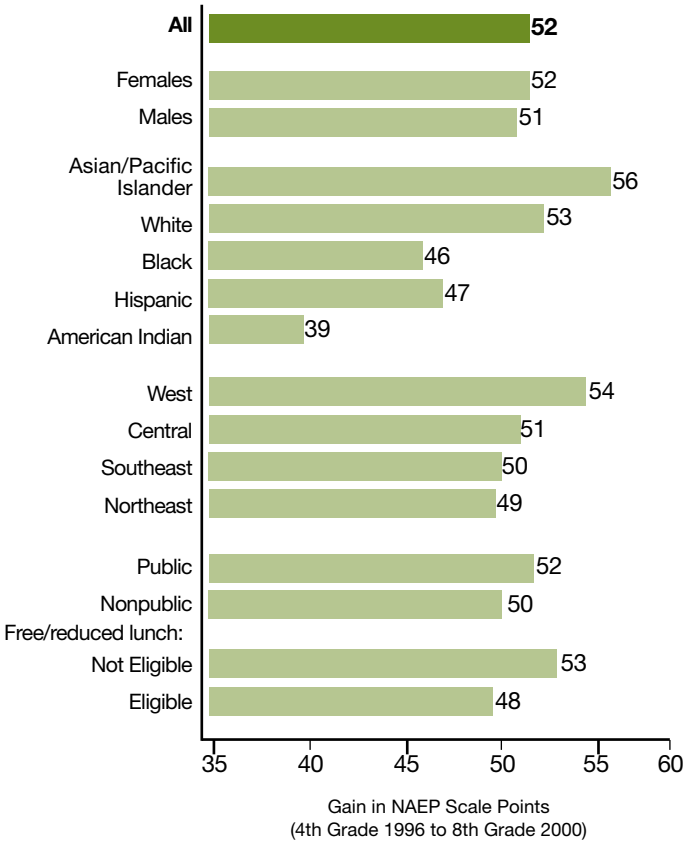
The comparable picture for mathematics, from 1996 to 2000, is shown in Figure 17. Here, Asian/Pacific Islander students grew the most, at 56 points. White students were next at 53, followed by Black and Hispanic students at 46 and 47, respectively. American Indian students were at the bottom with a gain of 39 points.[17]

Among the states, Minnesota led in eighth-grade scores and dropped to fifth in achievement gains from fourth to eighth grade.

For a full picture, it is important to examine both average score trends — as NAEP regularly reports — and trends in growth while students are in school. As this report has shown, the two measures tell us quite different things.

Figure 17

Cohort Growth in NAEP Mathematics, Grades 4 to 8, by Standard NAEP Reporting Groups



Source: Data from the National Assessment of Educational Progress analyzed by the ETS Policy Information Center.

[17] The only statistically significant difference is that White students showed more growth than Black and Hispanic students.

Scale scores do not convey the specific things that students can or cannot do. One way NAEP ascribes meaning to these numbers is through definitions of what represents achievement of *Advanced*, *Proficient*, and *Basic* performance.[18]

Another way to interpret NAEP results is to examine the specific kinds of questions students can answer and the problems they can solve at various points along the scale. Such information is reflected on something often called an "item map." Item maps provide concrete images of what students can and cannot do at particular score levels.

Figure 18 shows an item map for eighth-grade reading from the 2007 assessment. The map shows scores paired with representative items — tasks that students at the indicated scale score are likely to perform correctly and those lower on the scale are less likely to perform correctly. To the right are average scores for each racial/ethnic group and the average score for all students. Proficiency levels for the items are shown on the left. The average eighth-grader, with a score of 263, is likely to be able to identify causal relation between historical events, a skill that falls in the *Basic* proficiency level.

From Figure 18, we see that the average Asian American/Pacific Islander and White eighth-grader is likely to be able to describe the central problem faced by the main character. The average American Indian/Alaska Native and Hispanic eighth-grader is likely to be able to recognize information included by the author to persuade. The average Black eighth-grader is likely to be able to support opinion with text information or related prior knowledge. The average scores for all of these groups fall into the *Basic* proficiency level.

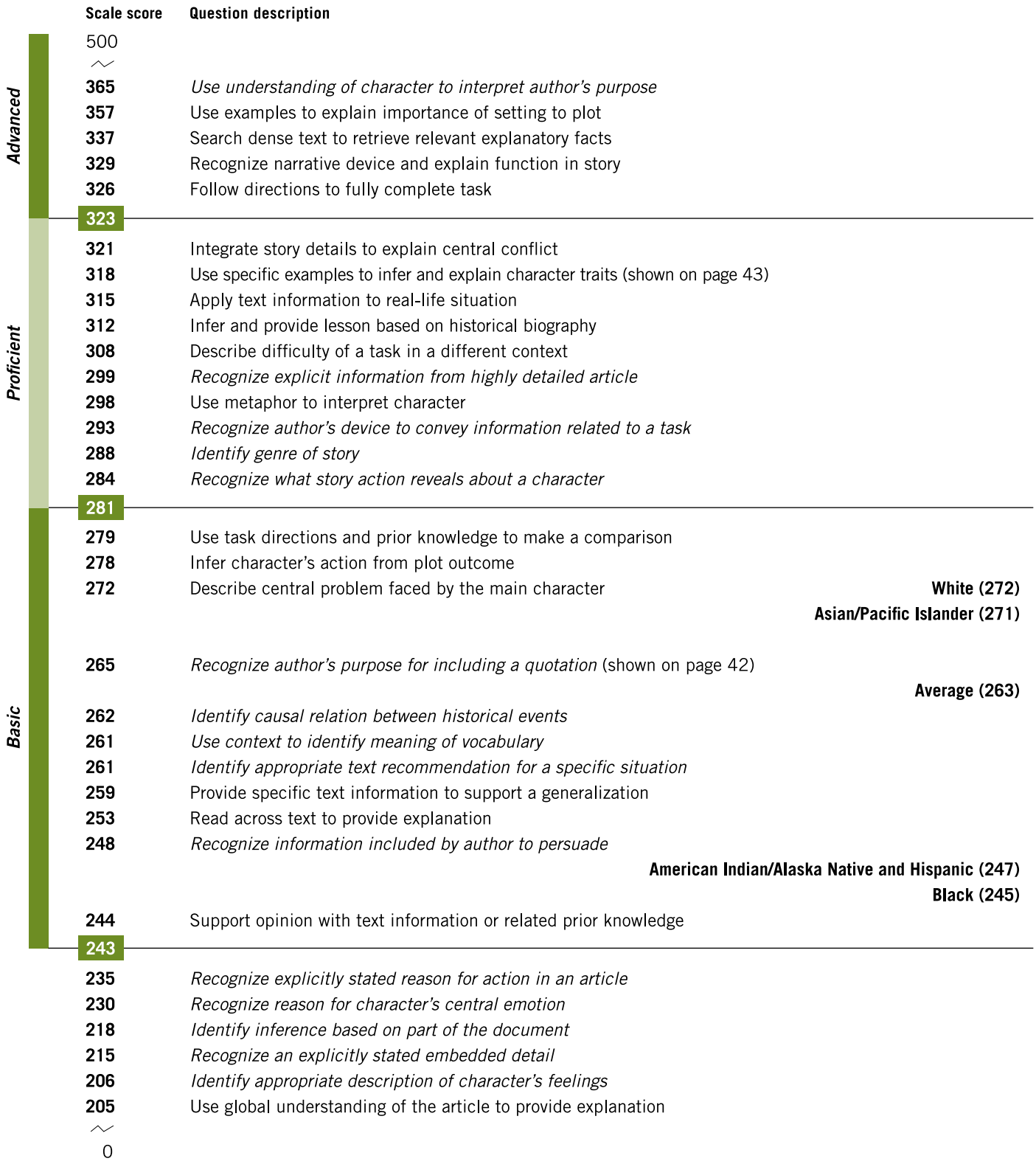
A mathematics item map is shown in Figure 19. It shows that the average eighth-grader can estimate time given a rate and a distance. The average Asian/Pacific Islander student can identify a relationship in a scatterplot, a task at the high end of the *Basic* proficiency level. The average Black eighth-grader performs just below the *Basic* level, likely to be able to evaluate an expression for a specific value.

These item maps are useful for understanding performance at different levels on the NAEP scale. If all achievement tests were reported in this way, communication among teachers, parents, students, and the public would likely improve.

[18] *Basic* denotes partial mastery of the knowledge and skills that are fundamental for proficient work at a given grade; *Proficient* represents solid academic performance and demonstrated competency over challenging subject matter; *Advanced* signifies superior performance. Detailed descriptions of the NAEP achievement levels for each grade and subject can be found on the website of the National Assessment Governing Board — http://www.nagb.org/pubs/pubs.html.

Figure 18

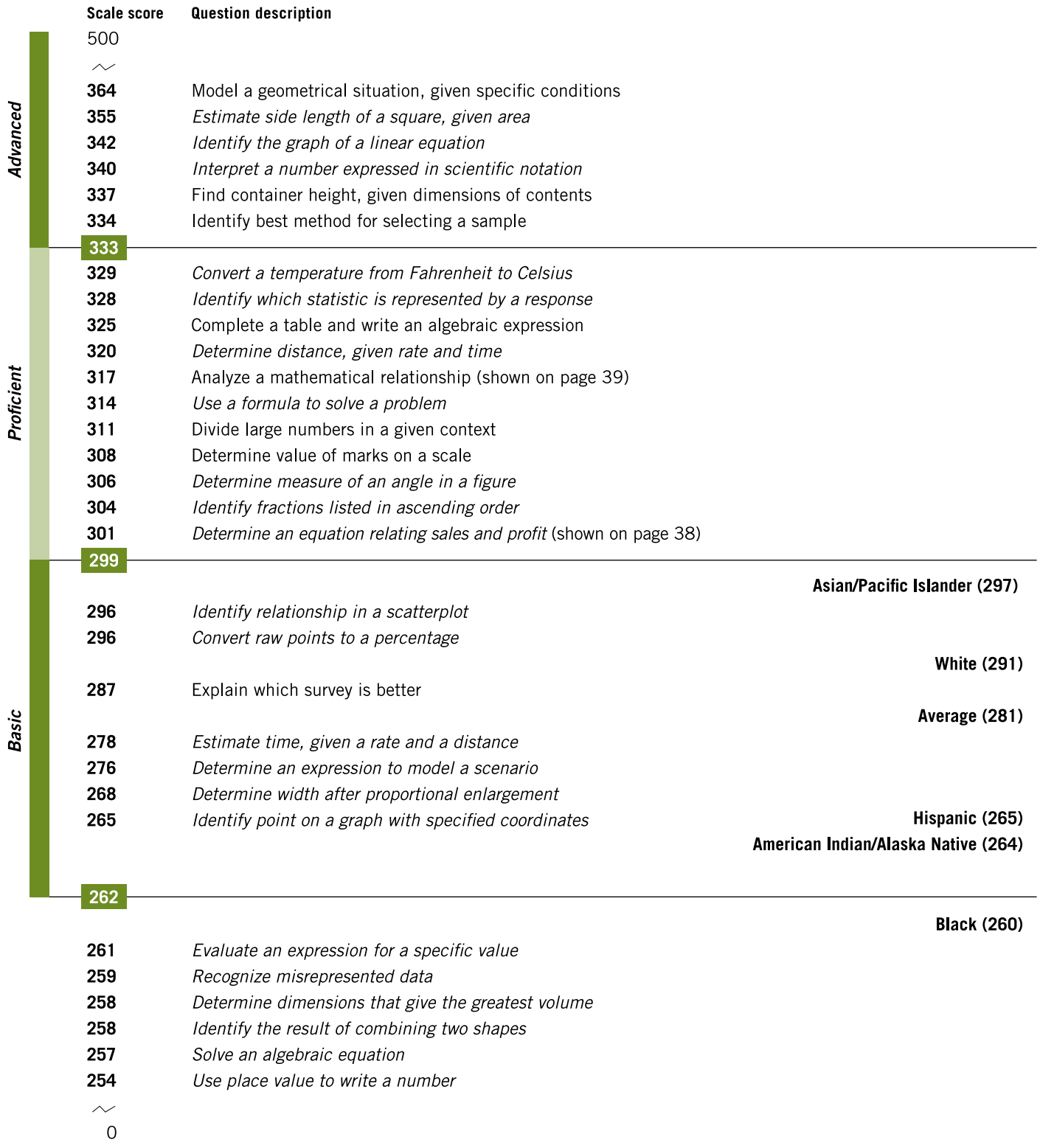
NAEP Grade 8 Reading Item Map



NOTE: Regular type denotes a constructed-response question. *Italic* type denotes a multiple-choice question. The position of a question on the scale represents the average scale score attained by students who had a 65 percent probability of successfully answering a constructed-response question, or a 74 percent probability of correctly answering a four-option multiple-choice question. For constructed-response questions, the question description represents students' performance rated as completely correct. Scale score ranges for reading achievement levels are referenced on the map.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Reading Assessment.

Figure 19

NAEP Grade 8 Mathematics Item Map



NOTE: Regular type denotes a constructed-response question. *Italic* type denotes a multiple-choice question. The position of a question on the scale represents the average scale score attained by students who had a 65 percent probability of successfully answering a constructed-response question, a 74 percent probability of correctly answering a four-option multiple-choice question, or a 72 percent probability of correctly answering a five-option multiple-choice question. For constructed-response questions, the question description represents students' performance rated as completely correct. Scale score ranges for mathematics achievement levels are referenced on the map.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Mathematics Assessment.

When we examine achievement measures, such as those used in this report, we see gaps: differences in average scores by race/ethnicity and by socioeconomic status. But while differences are clearly visible through the windows provided by these measures, one must examine each room in the house of the achievement gap to discover the conditions that create the gaps and then examine the classrooms of U.S. schools to learn what perpetuates them.

The challenge is to understand the experiences and conditions that affect achievement in critical ways, and then to find the measures and statistics that identify where the accompanying gaps are. Such knowledge is certainly attainable, at least to a satisfactory extent, if one has sufficient will, funds, and time.

A recent report from the ETS Policy Information Center details an informative exercise that was employed to help identify factors that contribute to achievement gaps.[18] First, the report synthesized research on the experiences and conditions associated with achievement (drawing on existing syntheses, as well as on individual research studies). After identifying factors in which the research community showed overall consensus, the report next located measures and statistics that showed whether differences existed in these experiences and conditions, on average, among people of different racial/ethnic and socio-economic backgrounds.

The first step illuminated 14 factors: six associated with schools, and eight associated with preschool and out-of-school experiences. These 14 factors are identified below, in bold face.

School

Teaching and Learning. The instructional infrastructure, including **rigor of the curriculum**, **teacher preparation**, **teacher experience and attendance**, **class size**, **availability of appropriate technology-assisted instruction**, and **school safety.**

Before and Beyond School

The Developmental Environment. The early experiences and conditions of life and living, including **weight at birth**; exposure to environmental hazards, such as **lead**; and **hunger and nutrition**.

The Home-Learning Connection. The support for learning in the home, including amount of time parents spent **reading to young children**, **amount of time children spent watching TV**, and **parent availability**.

The Community. The extent to which the community and its essential institutions support or hinder the efforts of families and schools.[20] Specifically, **student mobility** — how frequently children change schools — is related to socioeconomic status and can result in myriad problems in school.

The Home-School Connection. The two-way street of parents trying to be supportive of school efforts and schools reaching out to inform, encourage, and show receptivity to **parent participation,** which includes ensuring that children attend school regularly and encouraging children to do their homework.

One is likely to find intercorrelations both within and among these clusters, to varying degrees. For example, the developmental environment is likely to be closely related to community characteristics and support for education.

After identifying the 14 factors, the report focused on the statistics and measures available that could determine the related gaps by race/ethnicity and socioeconomic status. The results are in Table 7.

Measures were available for all 14 correlates by race/ethnicity, *and there were gaps in all 14 of them* — gaps that parallel measured school achievement. Comparable gaps were found by socioeconomic status for 11 of the 14 correlates, with data unavailable for two of them. The research indicates that if all these 14 correlates were controlled for, in statistical terms, the measured achievement gaps by race/ethnicity and socioeconomic status would all but disappear. (We cannot assume that all factors related to achievement have been found.)

Other research has produced similar findings. A 1993 issue of *ETS Policy Notes*, for example, identified five family factors strongly related to educational achievement. The positive factors were: having two parents in the home, reading more than two pages a day for school and homework, and having at least three types of reading material in the home; the

[19] Paul E. Barton, *Parsing the Achievement Gap: Baselines for Tracking Progress*, Policy Information Report, Policy Information Center, Educational Testing Service, 2003.
[20] Included here is the concept of social capital developed by James Coleman and Robert Putnam.

Table 7

Correlates of Achievement and Gaps

Correlates	Are There Gaps Between Minority and Majority Student Populations?	Are There Gaps Between Students from Low-Income Families and Higher-Income Families?
School:		
Rigor of Curriculum	Yes	Not Available
Teacher Preparation	Yes	Yes
Teacher Experience and Attendance	Yes	Yes
Class Size	Yes	No
Technology-assisted Instruction	Yes	Yes
School Safety	Yes	Yes
Before and Beyond School:		
Parent Participation	Yes	Yes
Student Mobility	Yes	Yes
Birthweight	Yes	Not Available
Lead Poisoning	Yes	Yes
Hunger and Nutrition	Yes	Yes
Reading to Young Children	Yes	Yes
Television Watching	Yes	Yes
Parent Availability	Yes	Yes

negative factors were student absenteeism and excessive television watching. *This analysis showed that 91 percent of the differences in mathematics test scores among the states were associated with these five factors*.[21]

A 2005 report by the ETS Policy Information Center examines the correlates of differences among states in high school completion rates.[22] This report found that socioeconomic status, the percentage of children under 18 living in two-parent families, and the percentage of students who did not change schools during the prior two years explained almost 60 percent of the variation in high school completion rates among states. When differences in the percentage of students of color were added to the estimate, the correlation was increased only slightly.

For the current report, calculations of differences among states and 10 large cities were made using eighth-grade reading scores on NAEP. But while NAEP is the best source for information on student achievement, it does not collect information on family income or if students live with both parents. To compensate, this report uses census data for family income and a proxy measure for one-parent families, informed by NAEP data on parent education.[23]

The results of multiple-regression analysis were these: family income alone accounted for 23 percent of the variation in NAEP achievement scores among the 50 states and 10 large cities. Entering the percent of students from one-parent families added 27 percent, for a total of 50 percent; and including the percent who were minority students added 26 percentage points, bringing the total percent to 76. This means that other student conditions and experiences play a substantial role in the differences among states and large cities in achievement — but that none is more substantial than the single-parent family factor.[24]

[21] *ETS Policy Notes* (1993), "Angles on Math Achievement," Policy Information Center, Educational Testing Service.
[22] Paul E. Barton, *One-Third of a Nation: Rising Dropout Rates and Declining Opportunities*, Policy Information Report, Policy Information Center, Educational Testing Service, February 2005.
[23] The measure used as a proxy for one-parent families was constructed from the data that NAEP collects on the education of students' parents in the student questionnaire. The percentage of students who said they did not know their mother's education level was subtracted from the percentage who did not know their father's education. The resulting percentage was assumed to be a rough measure of the students who were living with one parent — a measure that could be used to rank states and cities. The usefulness of this measure was confirmed by comparing it to the U.S. census measure of the percent of children who were not living in married-couple homes. This correlation was very high, at 0.85.
[24] See Appendix C for results of the regression analysis.

Securing a comprehensive picture of student achievement in the diverse and dynamic U.S. society requires tracking two developments: changes in academic achievement, which can be measured by NAEP scores, and changes in the racial/ethnic mix of our population. This report, along with many others, has shown that there are achievement gaps by race/ethnicity, and public policy is focused on reducing and eliminating the gaps. Much of this report is about those achievement gaps and how they are — or are not — changing. These two aspects of achievement can interrelate in several ways.

In one scenario, all subgroups could see score increases or decreases, resulting in respective increases or decreases in the average total score. This would not surprise anyone.

But consider other scenarios. What if scores for all — or most — subgroups increased, yet the average score remained the same, or even fell? Or what if scores in all subgroups fell, but the average score for all combined rose?

Either of these instances is possible, and statisticians have labeled the phenomenon "Simpson's Paradox." This phenomena occurs when a particular combination of changes in subgroup scores is seen alongside changes in the proportion of subgroups in the total student population.

For example, a subgroup with a considerably lower average score could have increased its average score, but the proportion of the group's population increased enough that it resulted in a decline in the overall average. One example of Simpson's Paradox: a mediocre student at a highly selective college transfers to a less-selective institution and raises the average achievement in both schools.

This section of the report will examine the combination of changing subgroup population shares among 17-year-old students and both changing NAEP reading scores from 1975 to 2004 and changing NAEP mathematics scores from 1978 to 2004. Also, the average score for all students in these grades will be compared with an average score that has been "standardized" for subgroup population share — that is, the score for 2004 will be adjusted to reflect the populations of racial/ethnic groups in 1975. This will effectively hold the proportions constant and indicate how much the changing proportions themselves affected the national average.[25]

Reading

For 17-year-olds, there was no change in the average reading score overall or in the average score for White students. However, Black, Hispanic, and "Other" 17-year-old students — a category that includes mostly Asian Americans — gained 23.2, 11.1, and 12.2 scale points, respectively. See Figure 20.

Large changes also occurred in the student population shares by race/ethnicity: the White share declined by 16.1 percentage points, the Hispanic population rose by 10.8 percentage points, and the Other category rose by 3.8 percentage points. The Black student population share did not change significantly.

The result was that the average score for all students remained unchanged, even though three of the four population groups gained in achievement and none showed a decline. What came into play was the changing demography, with the lower-scoring groups — although improving — limiting growth in the overall average score as their populations expanded. When standardized for race/ethnicity at the population shares that existed in 1975, however, the average score showed a gain of 3.8 scale points, somewhat more than the growth in the actual score.

Mathematics

In mathematics among 17-year-old students, average achievement rose by 6.3 scale points. However, all subgroups saw greater gains over the 26-year period: White students, 7.5 points; Black students, 16.9 points; Hispanic students, 13.1 points; and Other students (again, mostly Asian American), 6.7 points (see Figure 20).

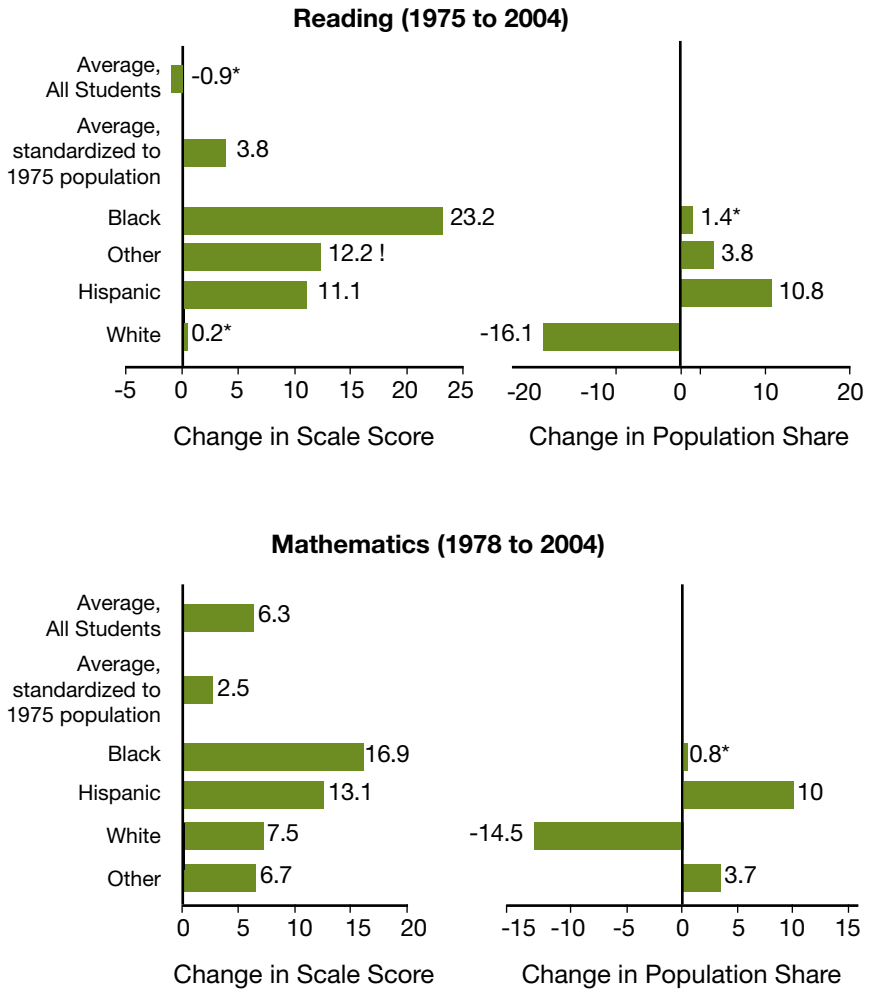
As with the reading scores, this seeming paradox is explained by the changing population shares of the different subgroups. The high-scoring White group declined substantially in population share, and the lower-scoring Hispanic group increased considerably in population share. The higher scoring Other group also increased its representation somewhat.

The average score, when standardized at the population distribution that existed in 1978, rose 2.5 points, somewhat less than the growth in the actual score.

[25] This standardization was also done for 9- and 13-year-olds with results similiar to the results for 17-year-olds.

Figure 20

Changes in NAEP Reading and Mathematics Scores and Population Shares, Age 17



*Not statistically significant
!High standard error; interpret with caution
Source: NAEP Long-Term Trend data, analyzed by ETS.

From time to time, Americans see a front-page news story about where the United States ranks in student achievement among other nations. This is particularly likely if the United States has been found to rank low in the comparisons.

Such international comparisons are a challenge to researchers and to the press because there are a lot of hurdles in the path of accurate measurement, comparison, and reporting. Experts differ on how to interpret the results and what the perceived differences in the comparative rankings mean for the U.S. economy of the future. And with so many surveys of so many students, in several different grades, and on several academic subjects, one is challenged to grasp the net findings.

A major step forward in comprehending the U.S. position internationally comes from a recent analysis carried out by Erling E. Boe and Sujie Shin at the University of Pennsylvania.[26] They aggregated surveys for 22 countries — surveys that spanned reading, mathematics, science, and civics. The International Monetary Fund classified the 22 countries studied as "industrialized," and the World Bank classified them as "High-Income OECD[27] Membership."

The aggregated surveys include:

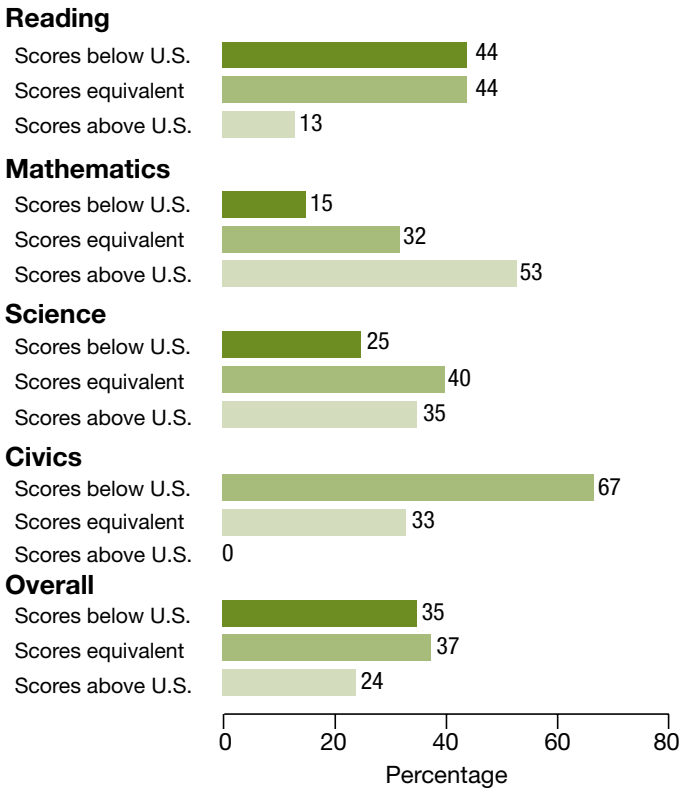
- The Reading Literacy Study (RLS), 1991 and 1996
- The Progress in International Reading Literacy Study (PIRLS), 1995 and its report in 1999
- The Third International Math and Science Study (TIMSS), 1995 and its report in 1999
- The Program for International Student Assessment (PISA), 2000 and 2001
- The Civic Education Study (CES), 1995

Figure 21 shows the overall results for each subject at all grades, for the United States, compared with the 21 nations.

- In *reading*, 13 percent of the nations scored above the United States, 44 percent were equivalent, and 44 percent scored below.
- In *mathematics*, 53 percent scored above the United States, 32 percent were equivalent, and 15 percent scored below.

Figure 21

Comparison of U.S. Performance with 21 Other Selected Countries, by Subject, All Grades



Source: Erling E. Boe and Sujie Shin, *Is the United States Winning or Losing the International Horse Race in Academic Achievement? Neither—It Is Running with Other Western G7 Nations*, Center for Research and Evaluation in Social Policy, Graduate School of Education, University of Pennsylvania, October 13, 2004.

- In *science*, 35 percent scored above the United States, 40 percent were equivalent, and 25 percent were below.
- In *civics*, none were above the United States, 33 percent were equivalent, and 67 percent were below.
- Aggregating *overall*, 24 percent were above the United States, 37 percent were equivalent, and 35 percent were below.

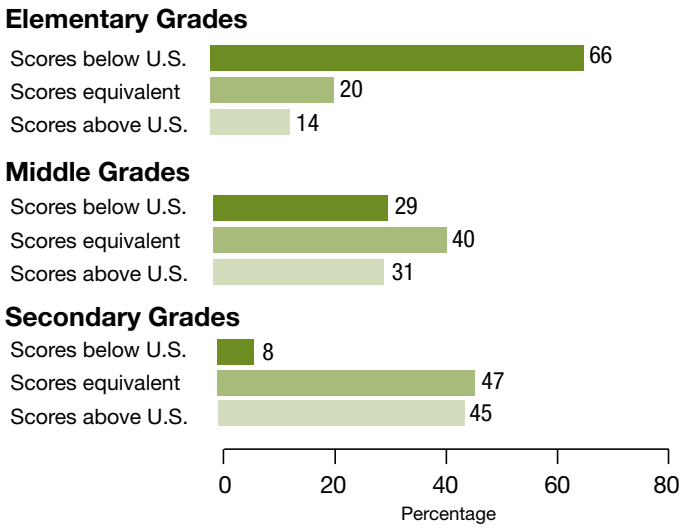
The results vary considerably by subject matter, with the United States clearly performing poorly in mathematics, well in civics, and about average in the rest. Combining all grades, however, conceals some important differences. Because of this, Boe and Shin also look at the grade levels separately; the results appear in Figure 22.

[26] Erling E. Boe and Sujie Shin, *Is the United States Winning or Losing the International Horse Race in Academic Achievement? Neither — It Is Running with Other Western G7 Nations*, Center for Research and Evaluation in Social Policy, Graduate School of Education, University of Pennsylvania, October 13, 2004.

[27] Organisation for Economic Co-operation and Development.

Figure 22

Comparison of U.S. Performance with 21 Other Selected Countries, by Grade Level



Source: Erling E. Boe and Sujie Shin, *Is the United States Winning or Losing the International Horse Race in Academic Achievement? Neither—It Is Running with Other Western G7 Nations*, Center for Research and Evaluation in Social Policy, Graduate School of Education, University of Pennsylvania, October 13, 2004.

The data from Boe and Shin show that the United States does best in the elementary grades, less well in the middle grades, and even worse in the secondary grades.

- At the *elementary* level, only 14 percent of nations scored above the United States, 20 percent were equivalent, and 66 percent were below the United States.
- In the *middle* grades, 31 percent were above the United States, 40 percent were equivalent, and 29 percent scored below.
- At the *secondary* level, 45 percent were above the United States, 47 percent were equivalent, and just 8 percent scored below.

These data indicate that U.S. achievement scores deteriorate considerably as grade levels climb. Boe and Shin do caution, however, that comparisons across secondary grades are not necessarily equal. The secondary grade scores include data from TIMSS (1995) based on "the final year of secondary school." But this final year varies by three to eight years depending on the country, and the longer students were in school, the better their scores were. So the United States, with just four years following eighth grade, is at a disadvantage; if the TIMSS data are removed, the United States does better but finishes still somewhat below average and sees scores decline as grade levels advance.

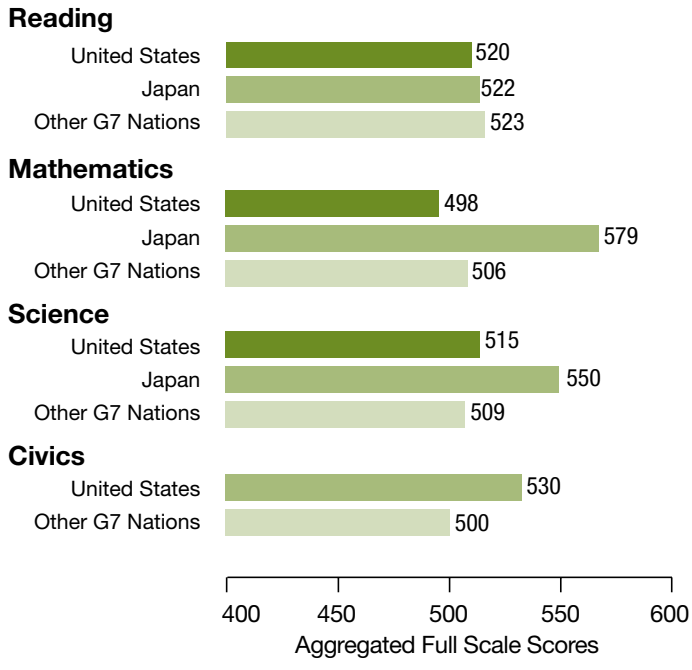
The authors sum it up this way:

> We conclude from the evidence reviewed above that U.S. students do not perform *poorly* in academic achievement compared with other industrialized nations. Instead, they perform better than average at the elementary grades and average at the middle and secondary grades across six international surveys and four subjects.[28]

Boe and Shin also specifically compare the United States with other G7 nations, other major economic competitors — Canada, France, Germany, Italy, Japan, and the United Kingdom. The researchers make comparisons at the middle and secondary levels, where the United States appeared least competitive among the 22 nations previously compared (see Figure 23).

Figure 23

Comparison of U.S. Performance with Japan and Other Western G7 Nations, by Subject, All Grades



Source: Erling E. Boe and Sujie Shin, *Is the United States Winning or Losing the International Horse Race in Academic Achievement? Neither—It Is Running with Other Western G7 Nations*, Center for Research and Evaluation in Social Policy, Graduate School of Education, University of Pennsylvania, October 13, 2004.

[28] Boe and Shin, 2004, p. 10.

What Boe and Shin found is that Japan is far ahead of all G7 nations in math and science. But the United States is comparable with the other five nations in reading, math, and science, and is ahead in civics.

One handicap the United States has in these kinds of surveys is having the highest level of socioeconomic diversification. A result of this diversity is the U.S. achievement gap, or lower averages in achievement among the Black and Hispanic populations. Closing this gap has high priority in the United States, as evidenced in the No Child Left Behind Act.

People will read different meanings into these international comparisons and will predict different consequences for the United States, based on average or below average performance. This composite picture is one way to illuminate where the United States stands in the international achievement arena.

Since Boe and Shin's composite, the results of a 2003 TIMSS mathematics assessment for grades four and eight and a 2003 PISA mathematics assessment for 15-year-olds became available. After the reports' initial release, the American Institutes for Research (AIR) re-analyzed the findings for the National Center for Education Statistics. The original report studied different countries for each grade, causing AIR to conclude that "because of the variability in the composition of countries participating in each assessment, these discussions have given an inaccurate impression that U.S. students' performance on PISA experienced a precipitous decline, compared with favorable U.S. rankings on TIMSS at grades four and eight."[29]

The AIR intended the new analysis to correct this mistaken impression and based its review on the common set of 12 countries that participated in all three assessments. The bottom line was as follows:

- At the fourth-grade level, seven countries were statistically above the United States and four were below.
- At the eighth-grade level, five countries were statistically above the United States, three were below, and three were not statistically different.
- When studying 15-year-olds, six countries were above the United States, three were below, and two were not statistically different.

These results are not much different from the composite of mathematics results by Boe and Shin, who found, for all levels combined, that 53 percent of countries scored above the United States. The 2003 results, in comparison, showed that 58 percent of countries scored above the United States at the fourth-grade level, 42 percent scored above the United States at the eighth-grade level, and 50 percent scored above the United States at age 15, for an average of 50 percent among the three.*

[29] Alan Ginsburg et al., *Reassessing U.S. International Mathematics Performance: New Findings from 2003 TIMSS and PISA*, American Institutes for Research, November 2005.

*As this report was going to press, new results from the *2006 Program for International Student Assessment* (PISA) were released showing that U.S. 15-year-olds ranked lower, on average, than their peers in 16 other countries in science, out of 30 developed nations taking part in the exam. U.S. students scored an average of 489, below the international average of 500 for industrialized nations. In math, which was tested in less depth on this PISA, U.S. students fared even worse, scoring 474, 24 points below the average for the 30 participating industrialized countries. In both science and math, U.S. students' performance was roughly the same as in 2003. Source: *Education Week*, "U.S. Students Fall Short in Math and Science," December 4, 2007.

There are no simple or short answers to these questions: What is the educational attainment of U.S. students at high school graduation, and at points along the way? How has our educational system been progressing regarding what students know and can do?

The critical national issue of educational achievement and inequality cannot be understood by looking at a couple of numbers. Alan Greenspan, who recently departed the Federal Reserve Board, likely looked at hundreds of economic indicators and is reported to have asked employees of statistical agencies for the latest esoteric numbers available. Improving educational achievement and eliminating achievement gaps are also complex issues, regardless of where one ranks education in comparison to the monetary system.

The importance of a vast reservoir of information that can be harnessed to pursue U.S. educational goals is not yet sufficiently appreciated. And although the quantity of data collected over the past several decades has increased considerably, public reports tend to focus on a particular survey. Drawing the surveys together helps to draw a more complete picture.

The development of cognitive ability begins at the very beginning of life. Thus, our knowledge about cognitive ability needs to start there, as does our investment in actions that create more equality in the conditions that promote its development. Preschool and out-of-school experiences and conditions are particularly critical and cannot afford to be neglected by policymakers and researchers.

Meanwhile, our knowledge base is reaping the benefits of the Department of Education's longitudinal study of kindergartners. The illuminated reality is that the large achievement gaps we see in fourth-grade NAEP data are already visible in kindergarten — as well as in the development of vocabulary in the first three years of life.

A broad understanding of achievement in the school years is often limited as the data reach the public through the media, which often looks for short, interesting nuggets that will attract attention.

The authors of this report believe that people must learn about more than the percentage of students who reach a particular score on a test. Bringing together a variety of ways to look at achievement in one place helps people understand how the top students are doing, what the status of students is at the bottom of the score distribution, as well as how well the average student is faring.

The data available on student achievement convey important information about what, in total, students know at the end of a school year and gauge progress by comparing previous classes' scores at the end of their school year — for example, this year's eighth-graders compared with other previous years' eighth-graders. We have added to that by using NAEP data to show how much students grew between fourth and eighth grade. We have also shown that state rankings using such methods can be quite different from rankings that result from scores representing the status of achievement at end-of-grade.

The United States has been undergoing, and will continue to experience, substantial changes in its demographic makeup. We gain clarity in our view of student achievement when we simultaneously look at these population changes and at each racial/ethnic group's performance level and trends.

Another important way to view achievement and inequality is by stepping back to the earliest years of life. Research has found the accompanying home and preschool experiences to be the precursors to academic performance in the classroom. This report has provided a brief summary of what is known about this important area.

Inequality also exists among nations, and the United States strives to claim the top position — and worries that low and faltering achievement scores could affect the country's world standing. But there is no straightforward way to compare U.S. students with students in other nations, and little unanimity exists in interpreting the meaning of international assessments. There is also the question of where the United States should rank — and what the country must do to rank higher. The composite picture presented in this report will not settle debates and disagreements, but perhaps it will serve as a helpful summary of international education comparisons over the last several years.

Overall, the authors believe it is necessary to bring together a wide array of information from many sources to gain a broad understanding of educational achievement in the United States.

****

Given the complex and nuanced views of achievement provided from different angles and at different stages of development, can readers glean some overall messages? The answer is yes.

First, differentiation in the crucial matter of language development begins between birth and the first three years of life, with babies internalizing what they see in their parents' faces, and what they hear from their parents' voices. The disparities in this early development are huge, and what little documentation exists on intervention efforts to remediate these deficits indicates that any successful effort would have to be intense and lengthy. Shortcuts to achieving an equal start do not look promising.

Not surprisingly, early differences show up clearly in kindergarten — as revealed by measures of developed verbal and mathematical abilities. Research shows, for example, that 44 percent of Asian American children can understand the beginning sounds of words at the beginning of kindergarten, compared with 20 percent of Black and Hispanic children. When viewed in terms of socioeconomic status, the percentages of kindergartners who can perform this task range from 51 percent down to 10 percent from the top to bottom quartiles. These data indicate that gross inequality in readiness to learn greets teachers at the onset of public education. And not unexpectedly, we see similar disparities among fourth- and eighth-graders regarding differences in average scores and in the percentages reaching what is defined as *Proficient* by NAEP.

A closer look shows what these differences mean in terms of the actual tasks students can perform. For example, while the average Asian American/Pacific Islander eighth-grader can identify a relationship in a scatterplot, the average Black eighth-grader can evaluate an expression for a specific value. It is important to convey what students know and can do in concrete terms such as this to convey the achievement gaps to the public and elected officials in more meaningful ways.

When researchers look beyond a state's average score or the percentage of students who are labeled as *Proficient* and instead identify the points on the score distribution at which changes are occurring, a different picture is presented to the public. The question raised is whether some U.S. schools are neglecting their top- and bottom-scoring students, and focusing on the group of students hovering just below the proficiency cut point. This is a question for policymakers and educators to consider before imposing sanctions on the basis of test scores.

And given what we know about the myriad factors that profoundly influence human development and educational achievement, we also need a clear view of the amount of learning that actually occurs during the school year. By following students in a sample, NAEP has provided us with a way to estimate changes in achievement from fourth to eighth grade. Ranking states by what students know in total at the end of a grade produces quite a different picture than one based on what students gained in knowledge over four years of schooling. States at the top in the first comparison do not necessarily maintain that position in the second. To receive this more robust account, several states and school districts are developing systems that allow individual student growth to be tracked.

The achievement gaps seen throughout the educational system also place the United States at risk of losing its competitive edge in the international arena. While top-performing students in the United States may be able to compete with the students of, say, Japan and South Korea, the combined U.S. population may fall short when compared with students in foreign countries with such relatively homogeneous student populations.

Attending to the achievement gap among U.S. students will go a long way toward narrowing the gap between the United States and its competitors. The fact that there are different international assessments, at different times, in different subjects, in different grades, and among different countries makes it difficult, however, to have a clear view of where the United States stands. Regardless, the available data indicate that U.S. students perform better than average in the elementary grades, and about average in the middle and secondary grades, with some variation among the subjects assessed.

Viewing student achievement against the backdrop of the nation's changing demographics provides another important insight. Research shows that racial/ethnic minority populations have been growing in the United States — and that these groups score lower, on average, on achievement tests.

With that in mind, a relevant question is what average scores on NAEP might be if the demographic makeup was the same as it was 25 or 30 years ago? We found that, after holding the population constant in terms of racial/ethnic composition, there would be little difference from the actual average scores attained recently.

Conveying a more complete story is necessary to judge U.S. success in schools and as a nation, as well as with U.S. states and communities. By looking in all the available windows on achievement and inequality, the United States can start to make strides in reducing the disparities that result in unequal achievement.

Appendix A: Changes in Four NAEP Reading Measures, Grade 8, 2002 to 2007

	Mean	Percent *Proficient*	Top Quartile	Bottom Quartile
Nation	-	-	-	-
Alabama	0	0	0	0
Arizona	0	0	0	-
Arkansas	0	0	0	0
California	0	0	0	0
Connecticut	0	0	0	0
Delaware	-	0	0	-
District of Columbia	0	0	+	0
Florida	0	0	0	0
Georgia	0	0	0	+
Hawaii	0	0	0	0
Idaho	0	0	-	0
Indiana	0	0	0	0
Kansas	0	0	-	0
Kentucky	-	-	0	-
Louisiana	0	0	0	0
Maine	0	0	0	+
Maryland	0	0	0	+
Massachusetts	0	0	+	0
Michigan	-	-	-	-
Mississippi	-	0	0	-
Missouri	-	0	0	-
Montana	0	0	0	0
Nebraska	-	0	0	-
Nevada	0	0	+	0
New Mexico	-	0	-	-
New York	0	0	+	-
North Carolina	-	0	-	-
North Dakota	0	0	-	+
Ohio	0	0	0	0
Oklahoma	-	0	0	-
Oregon	0	0	-	-
Pennsylvania	0	0	+	+
Rhode Island	-	0	0	-
South Carolina	0	0	0	0
Tennessee	0	0	-	0
Texas	0	0	0	0
Utah	0	0	-	0
Vermont	0	0	0	0
Virginia	0	0	0	0
Washington	-	0	-	0
West Virginia	-	-	-	-
Wyoming	0	0	0	0

+ = Statistically significant increase 0 = No significant change - = Statistically significant decline
Note: Includes states that participated in both assessments. No reading comparisons are available for Illinois, Minnesota, and Wisconsin because they did not meet reporting requirements in 2002.

Appendix B: Changes in Four NAEP Mathematics Measures, Grade 8, 2000 to 2007

	Mean	Percent *Proficient*	Top Quartile	Bottom Quartile
Nation	+	+	+	+
Alabama	0	0	0	+
Arizona	+	+	+	+
Arkansas	+	+	+	+
California	+	+	+	+
Connecticut	0	0	+	0
District of Columbia	+	+	+	+
DoDEA	+	+	+	+
Georgia	+	+	+	+
Hawaii	+	+	+	+
Idaho	+	+	+	+
Illinois	+	+	+	+
Indiana	+	+	+	0
Kansas	+	+	+	+
Kentucky	+	+	+	+
Louisiana	+	+	+	+
Maine	+	0	+	+
Maryland	+	+	+	+
Massachusetts	+	+	+	+
Michigan	0	0	+	0
Minnesota	+	0	+	+
Mississippi	+	+	+	+
Missouri	+	+	+	+
Montana	0	0	+	0
Nebraska	+	+	+	0
Nevada	+	+	+	+
New Mexico	+	+	+	+
New York	+	+	+	+
North Carolina	+	+	+	+
North Dakota	+	+	+	+
Ohio	+	+	+	0
Oklahoma	+	0	+	+
Oregon	0	0	+	0
Rhode Island	+	+	+	+
South Carolina	+	+	+	+
Tennessee	+	+	+	+
Texas	+	+	+	+
Utah	+	+	+	+
Vermont	+	+	+	+
Virginia	+	+	+	+
West Virginia	+	0	0	+
Wyoming	+	+	+	+

+ = Statistically significant increase 0 = No significant change - = Statistically significant decline
Note: Includes states that participated in both assessments.

Appendix C: Regression Analysis Used for Predicting Eighth-Grade NAEP Reading Proficiency

The dependent variable was the 2003 NAEP eighth-grade reading scores for 49 states and 10 major cities.* Median household income was used to control for socioeconomic status and was obtained from 1999 U.S. Census data. The next variable included is the percent of children who know their mothers' education level minus the percent who know their fathers' education level (these data are from the NAEP 2003 Reading Assessment and are used as a proxy measure for single-parent households). The final variable added was the percent of the population that was Black or Hispanic, as projected by the Census for 2003. An additional analysis was performed with the only difference being that instead of percent minority as the final variable added, the percent that lived in a different house one year ago was used as a proxy for mobility. This variable added no predictive power to the analysis and as such is not reported.

* The ten cities included in this analysis were: Atlanta; Boston; Charlotte; Chicago; Cleveland; Washington, D.C.; Houston; Los Angeles; New York; and San Diego. The one state not included was Alaska due to missing data for one variable. State data was not adjusted for states that also have a city represented — so, in effect, the urban areas have been given additional weight in the analysis.

This analysis was performed by Frederick Cline of ETS.

Model Summary

Variables included in Model	R	R^2	Adjusted R^2	Std. Error of the Estimate	Change Statistics			
					R^2 Change	F Change	df	Sig. F Change
1. Median household income (Source: 1999 U.S. Census)	.484	.234	.221	7.865	.234	17.456	57	< .001
2. + Percent who know mothers' education level minus the percent who know fathers' education level (Source: NAEP Eighth-Grade Reading Assessment, 2003)	.712	.507	.490	6.366	.273	31.028	56	< .001
3. + Percent of the population that is Black or Hispanic (Source: 2003 U.S. Census Projections)	.876	.768	.755	4.410	.260	61.685	55	< .001

Dependent Variable: NAEP Eighth-Grade Reading Scores for 2003

Variables		Unstandardized Coefficients		Standardized Coefficients			Correlations		
	Mean	B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part
(Constant)		254.152	4.546		55.904	.000			
Median household income	$47,993	.000	.000	.262	3.831	.000	.484	.459	.249
Percent who know mothers' education level minus the percent who know fathers' education level	8.39%	-.040	.210	-.018	- .191	.849	-.644	-.026	-.012
Percent of the population that is Black or Hispanic	24.51%	-.363	.046	-.750	- 7.854	.000	-.839	-.727	-.510

About ETS

ETS is a nonprofit institution with the mission to advance quality and equity in education by providing fair and valid assessments, research and related services for all people worldwide. In serving individuals, educational institutions and government agencies around the world, ETS customizes solutions to meet the need for teacher professional development products and services, classroom and end-of-course assessments, and research-based teaching and learning tools. Founded in 1947, ETS today develops, administers and scores more than 30 million tests annually in more than 180 countries, at over 9,000 locations worldwide.

Listening. Learning. Leading.®

www.ets.org