National Center for Education Statistics

The Nation's Report Card

Science 2000

U.S. Department of Education
Institute of Education Sciences

NCES 2003-453

What is The Nation's Report Card?

THE NATION'S REPORT CARD, the National Assessment of Educational Progress (NAEP), is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, history, geography, and other fields. By making objective information on student performance available to policymakers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement is collected under this program. NAEP guarantees the privacy of individual students and their families.

NAEP is conducted under the legislative authority of Congress. The Commissioner of the National Center for Education Statistics, under the U.S. Department of Education, is responsible by law for carrying out the NAEP project through competitive awards to qualified organizations. NAEP reports directly to the Commissioner, who is also responsible for providing continuing reviews, including validation studies and solicitation of public comment, on NAEP's conduct and usefulness.

In 1988, Congress established the National Assessment Governing Board (NAGB) to formulate policy guidelines for NAEP. The Board is responsible for selecting the subject areas to be assessed from among those included in the National Education Goals; for setting appropriate student performance levels; for developing assessment objectives and test specifications through a national consensus approach; for designing the assessment methodology; for developing guidelines for reporting and disseminating NAEP results; for developing standards and procedures for interstate, regional, and national comparisons; for determining the appropriateness of test items and ensuring they are free from bias; and for taking actions to improve the form and use of the National Assessment.

The National Assessment Governing Board

National Center for Education Statistics

The Nation's Report Card

Science 2000

Christine Y. O'Sullivan

Mary A. Lauko

Wendy S. Grigg

Jiahe Qian

Jinming Zhang

In collaboration with

Steven P. Isham **Satwinder Thind**

Youn-Hee Lim **Lois Worthington**

January 2003

U.S. Department of Education
Institute of Education Sciences

NCES 2003–453

U.S. Department of Education
Rod Paige
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Val Plisko
Associate Commissioner

January 2003

SUGGESTED CITATION
U.S. Department of Education. Institute of Education Sciences. National Center for Education Statistics. *The Nation's Report Card: Science 2000,* NCES 2003–453, by C.Y. O'Sullivan, M.A. Lauko, W. S. Grigg, J. Qian, and J. Zhang. Washington, DC: 2003.

FOR MORE INFORMATION
Content contact:
Taslima Rahman
202–502–7316

To obtain single copies of this report, limited number of copies available, or ordering information on other U.S. Department of Education products, call toll free 1–877–4ED-PUBS (877–433–7827), or write:

Education Publications Center (ED Pubs)
U.S. Department of Education
P.O. Box 1398
Jessup, MD 20794–1398

TTY/TDD 1–877–576–7734
FAX 301–470–1244

Online ordering via the Internet: *http://www.ed.gov/pubs/edpubs.html*
Copies also are available in alternate formats upon request.
This report also is available on the World Wide Web: *http://nces.ed.gov/pubsearch*

The work upon which this publication is based was performed for the National Center for Education Statistics by Educational Testing Service.

Table of Contents

The Nation's Report Card

Chapter 1: Tables and Figures

Chapter 2: Tables and Figures

Chapter 3: Tables and Figures

Chapter 4: Tables and Figures

Chapter 5: Tables and Figures

Executive Summary

The National Assessment of Educational Progress (NAEP) is the nation's only ongoing representative sample survey of student achievement in core subject areas. In 2000, NAEP conducted a national science assessment of fourth-, eighth-, and twelfth-grade students. State-level results were also collected at the fourth and eighth grades within participating states and jurisdictions.

Authorized by Congress and administered by the National Center for Education Statistics (NCES) in the U.S. Department of Education, NAEP regularly reports to the public on the educational progress of students in grades 4, 8, and 12. This report presents the results of the NAEP 2000 science assessment for the nation and the states. Results in 2000 are compared to results from the 1996 science assessment. Students' performance on the assessment is described in terms of average scores on a 0–300 scale for each grade and in terms of the percentages of students attaining three achievement levels: *Basic*, *Proficient*, and *Advanced*. The achievement levels are performance standards adopted by the National Assessment Governing Board (NAGB) as part of its statutory responsibilities and describe what students should know and be able to do. The Governing Board is an independent bipartisan group created by Congress in 1988 to set policy for NAEP.

As provided by law, the Deputy Commissioner of Education Statistics, upon review of a congressionally mandated evaluation of NAEP, determined that the achievement levels are to be considered developmental and should be interpreted and used with caution. However, both the Deputy Commissioner and NAGB believe these performance standards are useful for understanding trends in student achievement. They have been widely used by national and state officials as a common yardstick of academic performance.

In addition to providing average scores and achievement-level performance at the national level and state level, this report presents results for subgroups of students defined by various background and contextual characteristics. This report also contains results for a second sample at both the national and state levels—one in which testing accommodations were provided to students with special needs (i.e., students with disabilities or limited English proficient students).

The results presented in this report are based on representative samples of students for the nation and for participating states and jurisdictions. In the national sample, approximately 47,000 students from 2,100 schools were assessed. In the state samples, approximately 180,000 students from 7,500 schools were assessed. The national sample included students attending both public and nonpublic schools, while the state samples included only students attending public schools.

A summary of overall results from the 2000 NAEP science assessment is presented on the following pages. Differences between results from 1996 and 2000 or between groups of students are discussed only if they have been determined to be statistically significant.

Overall Science Results for the Nation, Regions, and States

Science Results for the Nation:

- Between 1996 and 2000, there was no statistically significant difference observed in the average science scores of fourth- or eighth-grade students. The average score of students in grade 12, however, declined from 150 in 1996 to 147 in 2000.
- In 2000, the percentage of students performing at or above *Proficient*—identified by NAGB as the level that all students should reach—was 29 percent at grade 4, 32 percent at grade 8, and 18 percent at grade 12. The percentage of eighth-graders at or above *Proficient* was higher in 2000 than in 1996. The percentage of twelfth-graders at or above *Basic* declined between 1996 and 2000.
- The 90th percentile score at grade 8 was higher in 2000 than in 1996, indicating improvement for the highest-performing eighth-graders. At grade 12, the 50th percentile score declined between 1996 and 2000, indicating a decline in the performance of middle-performing twelfth-graders.

Science Results for the Regions:

- In 2000, the average scores for fourth- and eighth-graders were higher in the Northeast and Central regions than in the Southeast and West. Among twelfth-graders, average scores were higher in the Northeast and Central regions than in the Southeast.
- Grade 12 students attending schools in the Central region had a lower average score in 2000 than in 1996.

Science Results for the States and Other Jurisdictions:

In the NAEP 2000 state-by-state assessment, results were reported for 39 states and 5 other jurisdictions that participated at grade 4, and 38 states and 4 other jurisdictions at grade 8. Only public schools participated in the state-by-state assessment.

At Grade 4:

- The top six states in 2000 were Iowa, Maine, Massachusetts, Montana, North Dakota, and Vermont. The average scores for these six states were higher than any other participating state but were not found to differ significantly from one another.
- Iowa, Maine, Massachusetts, Montana, and Vermont had percentages of fourth-graders at or above *Proficient* that were higher than the other participating states, but were not found to be significantly different from one another.

At Grade 8:

- The top 10 states and other jurisdictions in 2000 were Idaho, Maine, Massachusetts, Minnesota, Montana, North Dakota, Ohio, Vermont, and the Department of Defense domestic and overseas schools. The state of Montana, however, had an average eighth-grade score that was higher than any other participating state or jurisdiction.
- Between 1996 and 2000, eighth-graders' average scores increased in Missouri and at the Department of Defense domestic and overseas schools. (These results are based on multiple-comparison statistical significance testing procedures including all states or jurisdictions that participated in both years.)
- Massachusetts, Minnesota, Montana, and Ohio all had percentages of eighth-graders at or above *Proficient* that were higher than the percentages in other participating states, but were not found to differ significantly from one another.

National Science Results for Student Subgroups

In addition to overall results for the nation and for states and jurisdictions, NAEP reports on the performance of various subgroups of students. Observed differences between student subgroups in NAEP science performance most likely reflect a range of socioeconomic and educational factors not addressed in this report or by NAEP.

Gender

- In 2000, males had higher average scores than females at grades 4 and 8. The apparent gender difference at grade 12 was not statistically significant.
- Between 1996 and 2000, the average score for eighth-grade males increased, while the average score for twelfth-grade males decreased.
- Between 1996 and 2000, the average score gap favoring males over females widened by three points at grade 4 and by five points at grade 8.

Race/Ethnicity

- In 2000, the average scores of White students at all three grades were higher than those of their Black, Hispanic, or American Indian peers, and American Indian students scored higher on average than Black students.
- Between 1996 and 2000, average scores decreased for eighth-grade American Indian students and for twelfth-grade White students.
- Between 1996 and 2000, no significant difference was observed in the average score gap between White and Black students and between White and Hispanic students at any of the three grades.

Parents' Level of Education

- Generally, students in grades 8 and 12 who reported higher levels of parental education had higher average scores in 2000 than did their peers who reported lower levels of parental education. (Information about parental education was not collected at the fourth grade.)
- Between 1996 and 2000, average scores declined among twelfth-graders who reported that their parents' highest level of education was high school graduation and among those who reported that at least one parent had some education after high school.

Type of School

- At all three grades in 2000, students attending nonpublic schools had higher average scores than their peers attending public schools.
- Between 1996 and 2000, the average score for twelfth-grade public-school students decreased, while the average score for twelfth-grade nonpublic-school students increased.

Type of Location

- In 2000, fourth- and eighth-grade students attending schools in central city locations had lower average scores than their counterparts attending schools in urban fringe/large town or rural/small town locations. At grade 12, there was no statistically significant relationship between school location and students' average scores. (Results by type of location are not available from 1996.)

Free/Reduced-Price School Lunch Eligibility

- At all three grades in 2000, students eligible for the free/reduced-price school lunch program administered by the U.S. Department of Agriculture (USDA) had lower average scores than those who were not eligible. Free/reduced-price school lunches are intended for children at, near, or below the poverty line: eligibility is determined by the USDA's Income Eligibility Guidelines. (http://www.fns.usda.gov/cnd/IEGs&NAPs/IEGs.htm).
- Between 1996 and 2000, the average score of eighth-graders who were eligible for free/reduced-price school lunch decreased, while the average score of eighth-graders who were not eligible increased. Among twelfth-graders, the average score of students who were not eligible decreased between 1996 and 2000.

Becoming a More Inclusive NAEP

A second set of results from the NAEP 2000 science assessment includes the performance of special-needs students who were provided with testing accommodations. A similar set of results is available from 1996 at the national level only, allowing for comparisons between 1996 and 2000 national results based on administration procedures that permitted accommodations.

Science Results for the Nation:

- In 2000, the difference between "accommodations-permitted" and "accommodations-not-permitted" national average scores was not found to be statistically significant at grades 8 and 12. At grade 4, however, the "accommodations-permitted" average score was 2 points lower than the "accommodations-not-permitted" average score.[1]
- Between 1996 and 2000, the national average score for twelfth-graders declined when accommodations were not permitted and when accommodations were permitted.

Science Results for the States and Other Jurisdictions:

- In 2000, none of the apparent differences between "accommodations-permitted" and "accommodations-not-permitted" average scores were found to be statistically significant at either grade 4 or grade 8 for any of the participating states and jurisdictions. (These results are based on multiple-comparison statistical significance testing procedures including all states or jurisdictions that participated in 2000.)

---

[1] The effects of offering accommodations are examined in greater detail in two forthcoming reports:

Lutkus, A. D., & Mazzeo, J. *Including special-needs students in the NAEP 1998 reading assessment: Part I, comparison of overall results with and without accommodations.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (forthcoming).

Lutkus, A. D. *Including special-needs students in the NAEP 1998 reading assessment: Part II, results for students with disabilities and limited English proficient students.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (forthcoming).

School Contexts for Learning Science

NAEP collects information about the contexts for student learning by administering questionnaires to assessed students, their teachers, and school administrators. Using the student as the unit of analysis, NAEP examines the relationship between selected contextual variables drawn from these questionnaires and students' average scores on the science assessment. In interpreting these data, readers are reminded that the relationship between contextual variables and student performance is not necessarily causal. There are many factors that may play a role in student performance on NAEP.

Grade 4:

Computer Availability and Use

- In 2000, fourth-graders whose teachers reported that they used computers for science instruction scored higher, on average, than fourth-graders whose teachers reported that they did not.
- Between 1996 and 2000, the percentage of fourth-graders whose teachers reported using computers for science instruction increased from 47 to 57 percent.

Coursework

- In 2000, fourth-graders whose teachers reported spending a lot of time or some time on life science and Earth science had higher average scores than fourth-graders whose teachers reported spending only a little time on these domains.
- In 2000, 31 percent of fourth-grade students were taught by teachers who reported spending a lot of time on life science and Earth science, and 22 percent were taught by teachers who reported spending a lot of time on physical science.
- Between 1996 and 2000, the percentage of fourth-graders whose teachers reported spending a lot of time on Earth science increased from 19 to 31 percent.

Grade 8:

Computer Availability and Use

- In 2000, eighth-graders whose science teachers reported having their students use computers for simulations and modeling or for data analysis and other applications had higher average scores than eighth-graders whose science teachers reported not having students use computers in this manner.
- Between 1996 and 2000, the percentage of eighth-graders whose science teachers reported having their students use computers for data analysis and other applications or for word processing increased.

Coursework

- In 2000, 45 and 47 percent of eighth-graders were taught by teachers who reported spending a lot of time on Earth science and physical science, respectively. Twenty-one percent of eighth-graders were taught by teachers who reported spending a lot of time on life science.

Grade 12:

Computer Use

- In 2000, twelfth-graders who reported using computers to collect data or to analyze data in their science classes once a month or more had higher average scores than twelfth-graders who reported doing so less frequently.
- In 2000, twelfth-graders who reported never downloading data and related information from the Internet for their science classes had lower average scores than twelfth-graders who reported doing so at least sometimes.

Coursework

- Twelfth-graders who reported that they were currently taking a science course in 2000 scored higher, on average, than twelfth-graders who reported that they were not.
- According to twelfth-graders' reports in 2000 about the types of science courses they had taken since eighth-grade, approximately 74 percent had taken Earth science, 92 percent had taken biology, 70 percent had taken chemistry, and 36 percent had taken physics.
- Twelfth-grade students who reported in 2000 that they had taken or were currently enrolled in Advanced Placement (AP) biology, chemistry, or physics had higher average scores than twelfth-grade students who said they had not taken and were not enrolled in these AP courses.

The full set of results is available in an interactive database on the NAEP web site,

http://nces.ed.gov/nationsreportcard

Released test questions from the 1996 and 2000 science assessments and question-level performance data are also available on the web site.

1

NAEP 2000 Science Assessment

Introduction

National and international concern for students' achievement in science has been the impetus for several recent large-scale efforts to measure science knowledge and skills. For example, a repeat of the Third International Mathematics and Science Study (TIMSS-R) was conducted in 38 countries in 1999, 26 of which had also participated in 1995.[1] This assessment, conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA), was given to eighth-graders and measured students' knowledge and skills in the areas of mathematics and science. In 2000, the Program for International Student Assessment (PISA), organized by the Organization for Economic Cooperation and Development (OECD), was given to 15-year-olds in 32 countries for the first time.[2] This series of tests assessed reading literacy, mathematics literacy, and scientific literacy and was designed to measure the functional skills that students have acquired as they near the end of mandatory schooling. In addition to these assessments, in 2000

Chapter Focus

What is the NAEP science assessment?

How does the NAEP science assessment measure and report student progress?

Chapter Contents

[1] Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Gregory, K. D., Smith, T. A., Chrostowski, S. J., Garden, R. A., & O'Connor, K. M. (2000). *TIMSS 1999 international science report: Findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade.* Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College.

Gonzales, P., Calsyn, C., Jocelyn, L., Mak, K., Kastberg, D., Arafeh, S., Williams, T., & Tsen, W. (2000). *Pursuing excellence: Comparisons of international eighth-grade mathematics and science achievement from a U.S. perspective, 1995 and 1999* (NCES Publication No. 2001-028). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

[2] Lemke, M., Calsyn, C., Lippman, L., Jocelyn, L., Kastberg, D., Liu, Y., Roey, S., Williams, T., Kruger, T., & Bairu, G. (2000). *Outcomes of learning: Results from the 2000 program for international student assessment of 15-year-olds in reading, mathematics, and science literacy* (NCES Publication No. 2002–115). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

the National Assessment of Educational Progress (NAEP) administered assessments in science and mathematics to students at grades 4, 8, and 12, and in reading to students at grade 4. A number of states and other jurisdictions also took part in the science and mathematics assessments at grades 4 and 8. As with the TIMSS-R assessment, the NAEP assessments were designed to measure knowledge and skills in the various content domains.

The results of the TIMSS-R study and PISA study provide valuable information about the achievement of students in the United States vis-à-vis their counterparts throughout the world. For example, TIMSS-R showed that the science performance of eighth-graders in the U.S. was at the international average and no significant change was detected since the first TIMSS administration in 1995, and the PISA study showed that 15-year-olds in the U.S. performed at an average level in science literacy when compared to students in other countries.[3]

A voluntary Benchmarking Study was included as part of TIMSS 1999 that allowed the participating U.S. states and districts or consortia to assess the achievement of their students in an international context. Of the 13 states that participated in the study, all but 3 performed above the international average in science.[4]

The results of the NAEP 2000 science assessment provide important information about the performance of students in the nation, states, and other jurisdictions. This report discusses these results. It summarizes student achievement, compares results from the nation, states, and other jurisdictions, and discusses some of the many contextual variables collected during administration. In addition, the report also compares, where appropriate, results from the 1996 and 2000 science assessments. These results add to the body of information obtained from studies such as TIMSS-R and PISA about what students know and can do. The results also provide educators and policymakers with information that can be used to ascertain the well-being of science education in the U.S.

Overview of the 2000 National Assessment of Educational Progress (NAEP)

In 1969, NAEP was authorized by Congress to collect, analyze, and report reliable and valuable information about what American students know and can do in core subject areas. Since that time, in what has come to be referred to as the "long-term trend assessment," NAEP has assessed public- and nonpublic-school students who are 9, 13, and 17 years old. Since 1990, the

[3] Martin, M. O., Mullis, I.V. S., Gonzalez, E. J., Gregory, K. D., Smith, T. A., Chrostowski, S. J., Garden, R. A., & O'Connor, K. M. (2000). *TIMSS 1999 international science report: Findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade.* Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College.

Lemke, M., Calsyn, C., Lippman, L., Jocelyn, L., Kastberg, D., Liu, Y., Roey, S., Williams, T., Kruger, T., & Bairu, G. (2000). *Outcomes of learning: Results from the 2000 program for international student assessment of 15-year-olds in reading, mathematics, and science literacy* (NCES Publication No. 2002–115). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Gonzales, P., Calsyn, C., Jocelyn, L., Mak, K., Kastberg, D., Arafeh, S., Williams, T., & Tsen, W. (2000). *Pursuing excellence: Comparisons of international eighth-grade mathematics and science achievement from a U.S. perspective, 1995 and 1999* (NCES Publication No. 2001-028). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

[4] Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., O'Connor, K.M., Chrostowski, S.J., Gregory, K.D., Smith, T.A., & Garden, R.A. (2001). *Science benchmarking report: TIMSS 1999—eighth grade.* Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College.

more recently developed assessments, referred to as the main NAEP, have also assessed public- and nonpublic-school students in grades 4, 8, and 12. The results provided in this report from the 2000 science assessment are not comparable to those obtained from the 1999 science long-term trend assessment due to differences in the content of the two assessments, as well as different sampling and administration procedures.

All NAEP assessments are based on frameworks developed through a national consensus process. The 2000 NAEP science assessment was the second administration of an assessment based on *The NAEP Science Framework*.[5] In 1996 and 2000, the NAEP science assessment was administered to national samples of fourth-, eighth-, and twelfth-graders. The science assessment was also administered to samples of eighth-graders participating in the state-by-state assessment in 1996 and 2000 and to samples of fourth-graders participating in the state-by-state assessment in 2000. The legislation authorizing NAEP did not include state-by-state testing in grade 12.[6]

This report describes the results of the 2000 NAEP science assessment at grades 4, 8, and 12 and, where appropriate, compares results in 2000 to those in 1996. Comparisons across assessment years are possible because the assessments were developed under the same framework and share a common set of science questions, and because the populations of students in both assessments were sampled and assessed using comparable procedures.

The Science Assessment Framework

The NAEP Science Framework[7] provided the operational specifications and theoretical basis for developing NAEP science assessments in 1996 and 2000. It was developed in 1991 through a consensus process involving educators, policymakers, science teachers, representatives of the business community, assessment and curriculum experts, and members of the public. The project was managed by the Council of Chief State School Officers (CCSSO) under the auspices of the National Assessment Governing Board (NAGB).

The framework is organized along a content dimension and a cognitive dimension (knowing and doing). The content dimension is divided into three major fields of science: Earth, physical, and life. Science content pertaining to physics and chemistry is assessed within the field of physical science. The cognitive domain is divided into conceptual understanding, scientific investigation, and practical reasoning. Each question in the assessment is categorized by its content and cognitive domains. The framework also specifies two overarching categories—the nature of science and the organizing themes of science. Figure 1.1 summarizes the structure of the 1996 and 2000 assessments. The framework also specifies the percentage of assessment time to be devoted to each content and cognitive domain. A fuller description of the framework and a breakdown of the distribution of assessment time can be found in appendix A.

5 National Assessment Governing Board. (2000). *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*. Washington DC: Author. (Also available online at http://www.nagb.org/pubs/)

6 Public Law 100–297. (1988). National Assessment of Education Improvement Act (20 USC 1211).

7 National Assessment Governing Board. (2000). *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*. Washington DC: Author. (Also available online at http://www.nagb.org/pubs/)

Figure 1.1: Structure of the 2000 Assessment

Knowing and Doing	Fields of Science: Earth	Fields of Science: Physical	Fields of Science: Life
Conceptual Understanding			
Scientific Investigation			
Practical Reasoning			

SOURCE: National Assessment Governing Board. (2000). *Science Framework for the 1996 and 2000 National Assessment of Educational Progress.* Washington, DC: Author.

The Science Assessment Instruments

As the only federally mandated ongoing assessment of student science achievement on a national scale, it is imperative that NAEP reflect the framework and expert perspectives and opinions about science and its measurement. To that end, the assessment development process involves reviews by teachers and teacher educators as well as by state officials and measurement experts. All components of the assessment are evaluated for curricular relevance, developmental appropriateness, and fairness.

The 2000 science assessment booklets at grades 4, 8, and 12 consisted of two separately timed sections (i.e. blocks) of science questions that included both multiple-choice questions and constructed-response questions requiring students to create a written response. At the fourth grade, 20 minutes were allowed for each section of questions and at the eighth and twelfth grades, 30 minutes. In addition, one-half of the students in each school sample conducted a hands-on task and answered questions related to the task. For this, too, the time allotted was 20 minutes at grade 4 and 30 minutes at grades 8 and 12. It should be noted that students only took a portion of the assessment—two or three sections of the 14 sections that comprise the whole assessment at grades 4, 8, and 12. In addition to the science questions that students answered, they also responded to background questions that asked them to give information about themselves and their school experiences. For example, students were asked how much time they spent on homework, how often they used a computer, and what science subjects they were currently taking in school.

Additional information about the design of the 2000 science assessment is presented in appendix A.

Description of School and Student Samples

The NAEP 2000 science assessment was conducted nationally at grades 4, 8, and 12 and state-by-state at grades 4 and 8. The national assessment included representative samples of both public and nonpublic schools. The state-by-state assessments included only public schools. In total, 47,000 students from 2,100 schools were assessed in the national sample and 180,000 students from 7,500 schools in the state samples. Additional information about school and student samples is given in appendix A.

Jurisdictions including 40 states and 5 other jurisdictions participated in the state-by-state 2000 science assessment at grade 4 and 39 states and 5 other jurisdictions participated at grade 8. The 5 other jurisdictions that participated were American Samoa, the Department of Defense Domestic Dependent Elementary and Secondary Schools (DDESS), the overseas Department of Defense Dependents Schools (DoDDS), Guam, and the Virgin Islands. To ensure comparability across jurisdictions, NCES has established guidelines for school and student participation rates. Appendix A highlights these guidelines, and jurisdictions failing to meet them are noted in the tables and figures that present the state-by-state results.

Figure 1.2 lists the jurisdictions that participated in the 2000 science assessment and notes those jurisdictions that failed to meet one or more NCES-established participation rate guidelines for public schools. Results are not reported for the jurisdictions that failed to meet the initial school participation rate of 70 percent.

Figure 1.2 Participating jurisdictions in the NAEP 2000 state assessment program in science

Grade 4	Alabama	Louisiana	New York[2]	Virginia
	Arizona	Maine[2]	North Carolina	West Virginia
	Arkansas	Maryland	North Dakota	Wisconsin[1]
	California[2]	Massachusetts	Ohio[2]	Wyoming
	Connecticut	Michigan[2]	Oklahoma	American Samoa
	Georgia	Minnesota[2]	Oregon[2]	DDESS
	Hawaii	Mississippi	Rhode Island	DoDDS
	Idaho[2]	Missouri	South Carolina	Guam
	Illinois[2]	Montana[2]	Tennessee	Virgin Islands
	Indiana[2]	Nebraska	Texas	
	Iowa[2]	Nevada	Utah	
	Kentucky	New Mexico	Vermont[2]	
Grade 8	Alabama	Maine[2]	North Carolina	West Virginia
	Arizona[2]	Maryland	North Dakota	Wisconsin[1]
	Arkansas	Massachusetts	Ohio	Wyoming
	California[2]	Michigan[2]	Oklahoma	American Samoa
	Connecticut	Minnesota[2]	Oregon[2]	DDESS
	Georgia	Mississippi	Rhode Island	DoDDS
	Hawaii	Missouri	South Carolina	Guam
	Idaho[2]	Montana[2]	Tennessee	Virgin Islands[1]
	Illinois[2]	Nebraska	Texas	
	Indiana[2]	Nevada	Utah	
	Kentucky	New Mexico	Vermont[2]	
	Louisiana	New York[2]	Virginia	

[1] Failed to meet the initial school participation rate of 70 percent; results not reported.
[2] Failed to meet one or more participation rate guidelines; results reported with appropriate notation.
For more details on participation rate guidelines, see appendix A.
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools
DoDDS: Department of Defense Dependents School (Overseas)
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Two Sets of NAEP Results:

Accommodations Not Permitted and Accommodations Permitted

The NAEP assessments have always sought to include special-needs students—students with disabilities (SD) and limited English proficient students (LEP)—to the fullest degree possible. However, there have always been some special-needs students who were excluded from taking the NAEP assessment because they could not participate meaningfully in the assessment. Schools that participate in NAEP have been permitted to exclude some students who may have Individualized Education Programs (IEPs) or are receiving services under section 504 of the Rehabilitation act of 1973.[8] Similarly, schools have been permitted to exclude students they identify as being LEP. Schools are encouraged to make exclusion decisions in accordance with explicit criteria provided by NAEP.

[8] Section 504 of the Rehabilitation Act of 1973 is a civil rights law designed to prohibit discrimination on the basis of disability in programs and activities, including education, that receive federal financial assistance.

In order to move its assessments toward more inclusive samples, NAEP began to explore the use of accommodations, or alternate testing situations, with special-needs students in the 1996 science and mathematics assessments. This shift toward greater inclusiveness allowed NAEP to more closely approximate state and district testing policies that have increasingly offered testing accommodations to special-needs students. In 1996, the national NAEP sample was split so that some of the schools sampled were permitted to provide accommodations to special-needs students and the others were not. This split-sample design made it possible to study the effects on NAEP results of including special-needs students in the assessments under alternate testing conditions. A series of technical research papers has been published with the results of these comparisons.[9] Based on the outcomes of these technical analyses, the 1998 results of those NAEP assessments that used new test frameworks (writing and civics), and hence also began new trend lines, were reported for the first time with the inclusion of data from accommodated special-needs students.

This report includes two different sets of NAEP results based on the split-sample design:

- results based on a less inclusive sample that did not offer accommodations to special-needs students, and
- results based on a more inclusive sample that did offer accommodations (such as extended time and small-group administration) to special-needs students.

Although accommodated students make up a small proportion of the total weighted number of students assessed (see table A.9 in appendix A, for details), making accommodations available to special-needs students may change the overall assessment results in subtle ways. For example, some special-needs students who might have been tested without accommodations in previous assessment years might now receive accommodations, and, possibly, attain higher scores. Further, special-needs students who might have been excluded in previous years might now be included, but attain relatively low scores. The findings on results when accommodated special-needs students were included in the NAEP assessment are presented in chapter 4 of this report. In addition, appendix A contains a more detailed discussion of NAEP's intent to assess all students from the target population.

[9] Olsen, J. F., & Goldstein, A. A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments: A summary of recent progress* (NCES Publication No. 97–482). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Mazzeo, J., Carlson, J. E., Voelkl, K. E., & Lutkus, A. D. (1999). *Increasing the participation of special-needs students in NAEP: A report on 1996 research activities* (NCES Publication No. 2000–473). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Lutkus, A. D., & Mazzeo, J. *Including special-needs students in the NAEP 1998 reading assessment: Part I, comparison of overall results with and without accommodations.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (forthcoming).

Lutkus, A. D. *Including special-needs students in the NAEP 1998 reading assessment: Part II, results for students with disabilities and limited English proficient students.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (forthcoming).

Reporting the Assessment Results

The results of student performance on the NAEP science assessment are presented in two ways: one, as average scores on the NAEP composite science scale and two, as the percentage of students attaining NAEP science achievement levels. The average scale score represents students' performance on the assessment. The achievement levels reflect goals for student performance, and the percentage of students at each achievement level indicate the extent to which students are meeting those goals.

The composite scale at each grade ranges from 0 to 300. While the scale score ranges are identical, the scale was derived independently at each grade. Also scales were weighted differently at different grades in determining the overall scale. Therefore, average scale scores across grades cannot be compared. For example, equal scale scores on the grade 4 and grade 8 scales do not imply equal levels of science achievement. A full description of NAEP scales and scaling procedures can be found in the *NAEP 2000 Technical Report*.[10]

Achievement-level results are presented in terms of science achievement levels as authorized by the NAEP legislation and adopted by the National Assessment Governing Board (NAGB).[11] For each grade assessed, NAGB has adopted three achievement levels: *Basic*, *Proficient*, and *Advanced*. For reporting purposes, the achievement cut scores are placed on the science scale, resulting in four ranges: below *Basic*, *Basic*, *Proficient*, and *Advanced*.

The Setting of Achievement Levels

The 1988 NAEP legislation that created the National Assessment Governing Board directed the Board to identify "appropriate achievement goals.....for each subject area" that NAEP measures.[12] The 1994 NAEP reauthorization reaffirmed many of the Board's statutory responsibilities including "developing appropriate student performance standards for each age and grade in each subject area to be tested under the National Assessment."[13] In order to follow this directive and achieve the mandate of the 1988 statute to "improve the form and use of NAEP results," the Board undertook the development of student performance standards called "achievement levels." Since 1990 the Board has adopted achievement levels in mathematics, reading, U.S. history, world geography, science, writing, and civics.

The Board defined three levels for each grade: *Basic*, *Proficient*, and *Advanced*. The *Basic* level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at a given grade. The *Proficient* level represents solid

[10] National Center for Education Statistics. *NAEP 2000 technical report*. Washington, DC: Author (forthcoming).

[11] Public Law 100-297. (1988). National Assessment of Educational Progress Improvement Act (20USC 1211). Public Law 102-382. (1994). Improving America's Schools Act (20USC 9010).

[12] Public Law 100-297. (1988). National Assessment of Educational Progress Improvement Act (20USC 1211).

[13] Public Law 102-382. (1994). Improving America's Schools Act (20USC 9010).

academic performance for each grade assessed. Students reaching this level demonstrate competency over challenging subject matter. The *Advanced* level signifies superior performance at a given grade. Furthermore, for each grade, the levels are cumulative; that is, abilities achieved at the *Proficient* level presume mastery of abilities associated with the *Basic* level, and attainment of the *Advanced* level presumes mastery of both the *Basic* and *Proficient* levels.

Figure 1.3 presents the policy definitions of the achievement levels that apply across all grades and subject areas. Adopting three levels of achievement for each grade signals the importance of looking at more than one standard of performance. The Board believes, however, that all students should reach the *Proficient* level; the *Basic* level is not the desired goal, but rather represents partial mastery that is a step towards *Proficient*.

Figure 1.3

Achievement Levels

Policy definitions of the three achievement levels

Basic	This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.
Proficient	This level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.
Advanced	This level signifies superior performance.

SOURCE: National Assessment Governing Board.

The achievement levels in this report were arrived at somewhat differently from those adopted by the Board for other subject areas. A standard-setting process was carried out by ACT, Inc., under contract to the Board. ACT convened a cross section of educators and interested citizens across the nation and asked them to judge what students should know and be able to do relative to the body of content reflected in the NAEP framework for science. The achievement levels arrived at by this process were examined by the Board. In several cases, the levels appeared to be set either lower or higher than would be reasonable, resulting in too few or too many students placing at or above the *Basic, Proficient,* or *Advanced* levels. This belief was based on information about eighth-grade students from achievement levels adopted for other NAEP subjects, 1996 Advanced Placement (AP) results for twelfth-graders, and information about eighth-grade students from the Third International Mathematics and Science Study (TIMSS). The Board, therefore, adjusted the cut scores of some of the levels. Since the content descriptions developed by the ACT panelists no longer matched the cut scores adopted by the Board, a second panel of science educators and scientists was convened to develop new descriptions. These descriptions were based on student performance at each achievement level, and are a measure of what students know and can do. The new cut scores and content descriptions were adopted by the Board in 1996.[14]

Achievement-Level Descriptions for Each Grade

The achievement-level descriptions for grades 4, 8, and 12 are presented in figures 1.4 through 1.6. As noted previously, the achievement levels are cumulative. Therefore, students performing at the *Proficient* level also display the competencies associated with the *Basic* level, and students at the *Advanced* level also demonstrate the knowledge and skills associated with both the *Basic* and *Proficient* levels. For each achievement level listed in figures 1.4 through 1.6, the scale score that corresponds to the beginning of that level is shown in parentheses. For example, in figure 1.4 the scale score of 138 corresponds to the beginning of the grade 4 *Basic* level of achievement.

[14] Bourque, M. L., Champagne, A. B., & Crissman, S, (1997). *1996 science performance standards: Achievement results for the nation and the states.* Washington, DC: National Assessment Governing Board.

Figure 1.4 Grade 4	NAEP Science Achievement Levels
Basic (138)	**Students performing at the *Basic* level demonstrate some of the knowledge and reasoning required for understanding Earth, physical, and life sciences at a level appropriate to grade 4. For example, they can carry out simple investigations and read uncomplicated graphs and diagrams. Students at this level also show a beginning understanding of classification, simple relationships, and energy.** Fourth-grade students performing at the *Basic* level are able to follow simple procedures, manipulate simple materials, make observations, and record data. They are able to read simple graphs and diagrams and draw reasonable but limited conclusions based on data provided to them. These students can recognize appropriate experimental designs, although they are unable to justify their decisions. When presented with diagrams, students at this level can identify seasons; distinguish between day and night; and place the position of the Earth, sun, and planets. They are able to recognize major energy sources and simple energy changes. In addition, they show an understanding of the relationships between sound and vibrations. These students are able to identify organisms by physical characteristics and group organisms with similar physical features. They can also describe simple relationships among structure, function, habitat, life cycles, and different organisms.
Proficient (170)	**Students performing at the *Proficient* level demonstrate the knowledge and reasoning required for understanding of the Earth, physical and life sciences at a level appropriate to grade 4. For example, they understand concepts relating to the Earth's features, physical properties, structure, and function. In addition, students can formulate solutions to familiar problems as well as show a beginning awareness of issues associated with technology.** Fourth-grade students performing at the *Proficient* level are able to provide an explanation of day and night when given a diagram. They can recognize major features of the Earth's surface and the impact of natural forces. They are also able to recognize water in its various forms in the water cycle and can suggest ways to conserve it. These students recognize that various materials possess different properties that make them useful. Students at this level are able to explain how structure and function help living things survive. They have a beginning awareness of the benefits and challenges associated with technology and recognize some human effects on the environment. They can also make straightforward predictions and justify their position.
Advanced (205)	**Students performing at the *Advanced* level demonstrate a solid understanding of the Earth, physical, and life sciences as well as the ability to apply their understanding to practical situations at a level appropriate to grade 4. For example, they can perform and critique simple investigations, make connections from one or more of the sciences to predict or conclude, and apply fundamental concepts to practical applications.** Fourth-grade students performing at the *Advanced* level are able to combine information, data, and knowledge from one or more of the sciences to reach a conclusion or to make a valid prediction. They can also recognize, design and explain simple experimental procedures. Students at this level recognize nonrenewable sources of energy. They also recognize that light and sound travel at different speeds. These students understand some principles of ecology and are able to compare and contrast life cycles of various common organisms. In addition, they have a developmental awareness of the benefits and challenges associated with technology.

SOURCE: National Assessment Governing Board. (2000). *Science Framework for the 1996 and 2000 National Assessment of Educational Progress.* Washington, DC: Author.

Figure 1.5 Grade 8 — NAEP Science Achievement Levels

Basic (143)

Students performing at the *Basic* level demonstrate some of the knowledge and reasoning required for understanding of the Earth, physical, and life sciences at a level appropriate to grade 8. For example, they can carry out investigations and obtain information from graphs, diagrams, and tables. In addition, they demonstrate some understanding of concepts relating to the solar system and relative motion. Students at this level also have a beginning understanding of cause-and-effect relationships.

Eighth-grade students performing at the *Basic* level are able to observe, measure, collect, record, and compute data from investigations. They can read simple graphs and tables and are able to make simple data comparisons. These students are able to follow directions and use basic science equipment to perform simple experiments. In addition, they have an emerging ability to design experiments.

Students at this level have some awareness of causal relationships. They recognize the position of planets and their movement around the sun and know basic weather-related phenomena. These students can explain changes in position and motion such as the movement of a truck in relation to that of a car. They also have an emerging understanding of the interrelationships among plants, animals, and the environment.

Proficient (170)

Students performing at the *Proficient* level demonstrate much of the knowledge and many of the reasoning abilities essential for understanding of the Earth, physical, and life sciences at a level appropriate to grade 8. For example, students can interpret graphic information, design simple investigations, and explain such scientific concepts as energy transfer. Students at this level also show an awareness of environmental issues, especially those addressing energy and pollution.

Eighth-grade students performing at the *Proficient* level are able to create, interpret, and make predictions from charts, diagrams, and graphs based on information provided to them or from their own investigations. They have the ability to design an experiment and have an emerging understanding of variables and controls. These students are able to read and interpret geographic and topographic maps. In addition, they have an emerging ability to use and understand models, can partially formulate explanations of their understanding of scientific phenomena, and can design plans to solve problems.

Students at this level can begin to identify forms of energy and describe the role of energy transformation in living and nonliving systems. They have knowledge of organization, gravity, and motion within the solar system and can identify some factors that shape the surface of the Earth. These students have some understanding of properties of materials and have an emerging understanding of the particulate nature of matter, especially the effect of temperature on states of matter. They also know that light and sound travel at different speeds and can apply their knowledge of force, speed, and motion. These students demonstrate a developmental understanding of the flow of energy from the sun through living systems, especially plants. They know that organisms reproduce and that characteristics are inherited from previous generations. These students also understand that organisms are made up of cells and that cells have subcomponents with different functions. In addition, they are able to develop their own classification system based on physical characteristics. These students can list some effects of air and water pollution as well as demonstrate knowledge of the advantages and disadvantages of different energy sources in terms of how they affect the environment and the economy.

Figure 1.5 Grade 8 (continued)	NAEP Science Achievement Levels
Advanced (208)	**Students performing at the *Advanced* level demonstrate a solid understanding of the Earth, physical, and life sciences as well as the abilities required to apply their understanding in practical situations at a level appropriate to grade 8. For example, students can perform and critique the design of investigations, relate scientific concepts to each other, explain their reasoning, and discuss the impact of human activities on the environment.** Eighth-grade students performing at the *Advanced* level are able to provide an explanation for scientific results. They have a modest understanding of scale and are able to design a controlled experiment. These students have an understanding of models as representations of natural systems and can describe energy transfer in living and nonliving systems. Students at this level are able to understand that present physical clues, including fossils and geological formations, are indications that the Earth has not always been the same and that the present is a key to understanding the past. They have a solid knowledge of forces and motions within the solar system and an emerging understanding of atmospheric pressure. These students can recognize a wide range of physical and chemical properties of matter and some of their interactions and understand some of the properties of light and sound. Also, they can infer relationships between structure and function. These students know the difference between plant and animal cells and can apply their knowledge of food as a source of energy to a practical situation. In addition, they are able to explain the impact of human activities on the environment and the economy.

SOURCE: National Assessment Governing Board. (2000). *Science Framework for the 1996 and 2000 National Assessment of Educational Progress.* Washington, DC: Author.

Figure 1.6
Grade 12

NAEP Science Achievement Levels

Basic (146)	**Students performing at the *Basic* level demonstrate some knowledge and certain reasoning abilities required for understanding of the Earth, physical, and life sciences at a level appropriate to grade 12. In addition, they demonstrate knowledge of the themes of science (models, systems, and patterns of change) required for understanding the most basic relationships among the Earth, physical, and life sciences. They are able to conduct investigations, critique the design of investigations, and demonstrate a rudimentary understanding of the scientific principles.** Twelfth-grade students performing at the *Basic* level are able to select and use appropriate simple laboratory equipment and write down simple procedures that others can follow. They also have a developmental ability to design complex experiments. These students are able to make classifications based on definitions such as physical properties and characteristics. Students at this level demonstrate a rudimentary understanding of basic models and can also identify some parts of physical and biological systems. They are also able to identify some patterns in nature and rates of change over time. These students have the ability to identify basic scientific facts and terminology and have a rudimentary understanding of the scientific principles underlying such phenomena as volcanic activity, disease transmission, and energy transformation. In addition, they have some familiarity with the application of technology.
Proficient (178)	**Students performing at the *Proficient* level demonstrate the knowledge and reasoning abilities required for understanding of the Earth, physical, and life sciences at a level appropriate to grade 12. In addition, they demonstrate knowledge of the themes of science (models, systems, and patterns of change) required for understanding how these themes illustrate essential relationships among the Earth, physical, and life sciences. They are able to analyze data and apply scientific principles to everyday situations.** Twelfth-grade students performing at the *Proficient* level are able to demonstrate a working ability to design and conduct scientific investigations. They are able to analyze data in various forms and utilize information to provide explanations and to draw reasonable conclusions. Students at this level have a developmental understanding of both physical and conceptual models and are able to compare various models. They recognize some inputs and outputs, causes and effects, and interactions of a system. In addition, they can correlate structure to function for the parts of a system that they can identify. These students also recognize that rate of change depends on initial conditions and other factors. They are able to apply scientific concepts and principles to practical applications and solutions for problems in the real world and show developmental understanding of technology, its uses, and its applications.

Figure 1.6 Grade 12 (continued)	NAEP Science Achievement Levels
Advanced (210)	**Students performing at the *Advanced* level demonstrate the knowledge and reasoning abilities required for a solid understanding of the Earth, physical, and life sciences at a level appropriate to grade 12. In addition, they demonstrate knowledge of the themes of science (models, systems, and patterns of change) required for integrating knowledge and understanding of scientific principles from Earth, physical, and life sciences. Students can design investigations that answer questions about real-world situations and use their reasoning abilities to make predictions.** Twelfth-grade students performing at the *Advanced* level are able to design scientific investigations to solve complex, real-world situations. They can integrate, interpolate, and extrapolate information embedded in data to draw well-formulated explanations and conclusions. They are also able to use complex reasoning skills to apply scientific knowledge to make predictions based on conditions, variables, and interactions. Students at this level recognize the inherent strengths and limitations of models and can revise models based on additional information. They are able to recognize cause-and-effect relationships within systems and can utilize this knowledge to make reasonable predictions of future events. These students are able to recognize that patterns can be constant, exponential, or irregular and can apply this recognition to make predictions. They can also design a technological solution for a given problem.

SOURCE: National Assessment Governing Board. (2000). *Science Framework for the 1996 and 2000 National Assessment of Educational Progress.* Washington, DC: Author.

The Developmental Status of Achievement Levels

The 1994 NAEP reauthorization law requires that the achievement levels be used on a developmental basis until the Commissioner of Education Statistics determines that the achievement levels are 'reasonable, valid, and informative to the public." [15] Until the determination is made, the law requires the Commissioner and NAGB to state clearly the developmental status of the achievement levels in all NAEP reports.

In 1993, the first of several congressionally mandated evaluations of the achievement-level-setting process concluded that the procedures used to set the achievement levels were flawed and that the percentage of students at or above any particular achievement level cutpoint may be underestimated.[16] Others have asserted that the weight of the empirical evidence does not support such conclusions.[17]

[15] Improving America's Schools Act of 1994 (20 USC 9010) requires that the Commissioner base his determination on a congressionally mandated evaluation by one or more nationally recognized evaluation organizations.

[16] United States General Accounting Office. (1993). *Education achievement standards: NAGB's approach yields misleading interpretations*, U.S. General Accounting Office Report to Congressional Requestors. Washington, DC: Author.

National Academy of Education. (1993). *Setting performance standards for achievement: A report of the National Academy of Education panel on the evaluations of the NAEP trial state assessment: An evaluation of the 1992 achievement levels*. Stanford, CA: Author.

[17] Cizek, G (1993). *Reactions to National Academy of Education report.* Washington, DC: National Assessment Governing Board.

Kane, M. (1993). *Comments on the NAE evaluation of the NAGB achievement levels.* Washington, DC: National Assessment Governing Board.

The most recent congressionally mandated evaluation conducted by the National Academy of Sciences (NAS) relied on prior studies of achievement levels, rather than carrying out new evaluations, on the grounds that the process has not changed substantially since the initial problems were identified. Instead, the NAS Panel studied the development of the 1996 science achievement levels. The NAS Panel basically concurred with earlier congressionally mandated studies. The Panel concluded that "NAEP's current achievement level setting procedures remain fundamentally flawed. The judgment tasks are difficult and confusing; raters' judgments of different item types are internally inconsistent; appropriate validity evidence for the cut scores is lacking; and the process has produced unreasonable results." [18]

A proven alternative to the current process has not yet been identified. The Deputy Commissioner of Education Statistics and the Board continue to call on the research community to assist in finding ways to improve standard setting for reporting NAEP results. The NAS Panel accepted the continuing use of achievement levels in reporting NAEP results, until such time as better procedures can be developed. Specifically, the NAS Panel concluded that "...tracking changes in the percentages of students performing at or above those cut scores (or, in fact, any selected cut scores) can be of use in describing changes in student performance over time."[19] The National Assessment Governing Board urges all who are concerned about student performance levels to recognize that the use of these achievement levels is a developing process and is subject to various interpretations. The Board and the Deputy Commissioner believe that the achievement levels are useful for reporting trends in the educational achievement of students in the United States. However, based on the congressionally mandated evaluations so far, the Deputy Commissioner agrees with the National Academy's recommendation that caution needs to be exercised in the use of the current achievement levels. Therefore, the Deputy Commissioner concludes that these achievement levels should continue to be considered developmental and continue to be interpreted and used with caution.

[18] Pellegrino, J.W., Jones, L.R., & Mitchell, K.J. (Eds.). (1999). *Grading the nation's report card:Evaluating NAEP and transforming the assessment of educational progress.* Committee on the Evaluation of National Assessments of Educational Progress, Board on Testing and Assessment, Commission on Behavioral and Social Sciences and Education, National Research Council. (pp.182). Washington, DC: National Academy Press.

[19] Ibid., 176.

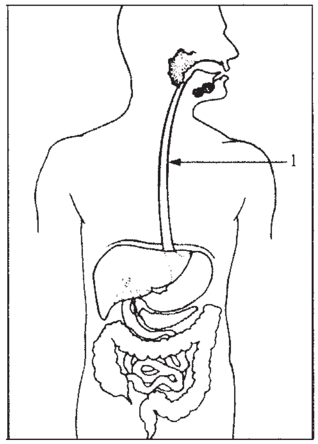
Sample Assessment Questions

Three blocks of questions at each grade were released to the public following the administration of the NAEP 2000 science assessment and three blocks at each grade were released following the 1996 administration. All these questions can be found on the NAEP web site.[20] The questions released from the 2000 assessment were also administered in the 1996 science assessment. Results for nine of the released questions, three from each of grades 4, 8, and 12, are presented in tables 1.1 through 1.9. These questions illustrate the types of questions included in the assessment.

The first three sample questions were administered at grade 4. Sample question 1 is a life science question that asked students to recognize the function of the esophagus. Sample question 2 is an Earth science question that asked students to explain why the Earth never runs out of rain. Sample question 3 is a physical science question that required students to explain how they could find out which of the three differently shaped bottles would hold the most water.

[20] http://nces.ed.gov/nationsreportcard

Grade 4 Sample Question 1:



Look at the picture above, which shows some of the organs that can be found inside the human body. What is the main job of the organ labeled 1 ?

Ⓐ Carrying air

● Carrying food

Ⓒ Carrying blood

Ⓓ Carrying messages from the brain

Table 1.1 Sample Question 1 Results (Multiple-Choice)

Overall percentage correct and percentages correct within each achievement-level range: 2000

Grade 4	Percentage correct within achievement-level intervals		
Overall percentage correct †	*Basic* 138–169*	*Proficient* 170–204*	*Advanced* 205 and above*
55	55	75	90

† Includes fourth-grade students who were below the *Basic* level.
*NAEP Science composite scale range.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Grade 4 Sample Question 2:

Think about where rain comes from and explain why the Earth never runs out of rain.

Responses to this question were scored according to a three-level rubric as Unsatisfactory/Incorrect, Partial, or Complete.

Sample "Complete" Response:

When we get rain it evaprates and rains again.

Earth never runs out of rain because the water from the ocean keeps evaporating.

Table 1.2 Sample Question 2 Results (Short Constructed-Response)

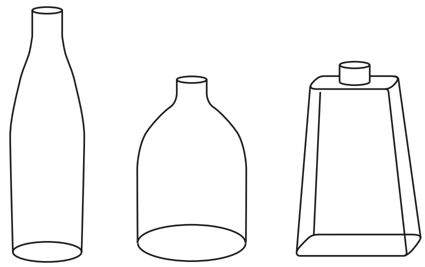
Overall percentage "Complete" and percentages "Complete" within each achievement-level range: 2000

Grade 4	Percentage "Complete" within achievement-level intervals		
Overall percentage "Complete" †	*Basic* 138–169*	*Proficient* 170–204*	*Advanced* 205 and above*
28	26	45	65

†Includes fourth-grade students who were below the *Basic* level.
*NAEP Science composite scale range.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Grade 4 Sample Question 3:

You are going to the park on a hot day and need to take some water with you. You have three different bottles, as shown in the picture below. You want to choose the bottle that will hold the most water. Explain how you can find out which bottle holds the most water.



Responses to this question were scored according to a three-level rubric as Unsatisfactory/Incorrect, Partial, or Complete.

Sample "Complete" Response:

You can get a couple cups that are the same size, poor water in the cups and poor the water in the jugs and wich ever one takes the most cups can fit more water.

Table 1.3 Sample Question 3 Results (Short Constructed-Response)

Overall percentage "Complete" and percentages "Complete" within each achievement-level range: 2000

Grade 4	Percentage "Complete" within achievement-level intervals		
Overall percentage "Complete" †	*Basic* 138–169*	*Proficient* 170–204*	*Advanced* 205 and above*
5	4	10	23

† Includes fourth-grade students who were below the *Basic* level.
*NAEP Science composite scale range.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

The next three sample questions were administered at grade 8. Sample question 4 is a physical science question that asked students to recognize that a magnet would not be helpful in separating a mixture of sand and salt. Sample question 5 assessed the domain of Earth science. It required students to state what they thought caused a monument to crumble and how to prevent further damage to the stone. Sample question 6 asked students to place 8 animals into two groups based on a physical characteristic. They were also asked to name a second physical characteristic they could have used. This was a life science question.

Grade 8 Sample Question 4:

All of the following would be helpful in separating a mixture of sand and salt EXCEPT

- (A, filled in) a magnet
- (B) a glass cup
- (C) a filter paper and funnel
- (D) water

Table 1.4 Sample Question 4 Results (Multiple-Choice)

Overall percentage correct and percentages correct within each achievement-level range: 2000

Grade 8	Percentage correct within achievement-level intervals		
Overall percentage correct †	*Basic* 143–169*	*Proficient* 170–207*	*Advanced* 208 and above*
59	59	71	81

†Includes eighth-grade students who were below the *Basic* level.
*NAEP Science composite scale range.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Grade 8 Sample Question 5:

Cleopatra's Needle is a large stone monument that stood in an Egyptian desert for thousands of years. Then it was moved to New York City's Central Park. After only a few years, its surface began crumbling.

What probably caused this crumbling?

____

____

New York City wants to keep Cleopatra's Needle in the same location in Central Park. How can the city prevent further damage to the stone?

____

____

Responses to this question were scored according to a three-level rubric as Unsatisfactory/Incorrect, Partial, or Complete.

Sample "Complete" Response:

What probably caused this crumbling?

Because of the polution and acid rian.

New York City wants to keep Cleopatra's Needle in the same location in Central Park. How can the city prevent further damage to the stone?

They could put roof over it or something to protect it from the rain

Table 1.5 Sample Question 5 Results (Short Constructed-Response)

Overall percentage "Complete" and percentages "Complete" within each achievement-level range: 2000

Grade 8	Percentage "Complete" within achievement-level intervals		
Overall percentage "Complete" †	*Basic* 143–169*	*Proficient* 170–207*	*Advanced* 208 and above*
28	28	47	71

†Includes eighth-grade students who were below the *Basic* level.
*NAEP Science composite scale range.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Grade 8 Sample Question 6:

Classify each of the eight living things listed below into one of two groups according to an important physical characteristic.

Gorilla
Parrot
Snake
Earthworm
Jellyfish
Sponge
Fish
Fly

Group 1 | *Group 2*

What physical characteristic did you use in your classification?

____

____

Name a different physical characteristic that you could have used.

____

____

Responses to this question were scored according to a four-level rubric as Unsatisfactory/Incorrect, Partial, Essential, or Complete.

Sample "Complete" Response:

Classify each of the eight living things listed below into one of two groups according to an important physical characteristic.

Group 1		Group 2
Gorilla	Gorilla	Snake
Parrot	Parrot	fish
Sponge	Snake	
earthworm	Earthworm	
fly	Jellyfish	
	Sponge	
	Fish	
	Fly	

What physical characteristic did you use in your classification?

I characterized them by having scales & not having scales

Name a different physical characteristic that you could have used.

having wings and not having wings

Table 1.6 Sample Question 6 Results (Extended Constructed-Response)

Overall percentage "Essential" and percentages "Essential" within each achievement-level range: 2000

Grade 8	Percentage "Essential" or better within achievement-level intervals		
Overall percentage "Essential" or better †	*Basic* 143–169*	*Proficient* 170–207*	*Advanced* 208 and above*
24	23	40	67

†Includes eighth-grade students who were below the *Basic* level.
*NAEP Science composite scale range.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

The last three sample questions were administered at grade 12. The earth science question shown in sample 7 required students to decide which of four statements most likely explained the observation that the Sun appears to be slightly larger in January than in July. Sample question 8 required students to describe a procedure for determining the density of a ring. It was classified as a physical science question. Sample 9, a life science question, was one of a set of 8 questions that probed students' understanding of genetics and genetic engineering. Students were first asked to read an article about the use of viruses in genetic engineering and then asked to use the information in the article plus their own knowledge of genetics to answer the series of questions. The first question in the set asked students to state what a gene is, what it is made of, and its function.

Grade 12 Sample Question 7:

As observed with special instruments from Earth, the Sun appears in the sky to be slightly larger in January than in July. Which of the following accounts for this observation?

- (A) [filled] The Earth moves in an orbit that is not circular but is closer to the Sun in January than in July.
- (B) The diameter of the Earth is not constant, but bulges slightly at the Equator and contracts slightly during the winter.
- (C) The Earth's orbit is not in the same plane as the orbits of the other planets.
- (D) The axis of rotation of the Earth is not perpendicular to the plane of its orbit but instead is tilted at an angle.

Table 1.7 Sample Question 7 Results (Multiple-Choice)

Overall percentage correct and percentages correct within each achievement-level range: 2000

Grade 12	Percentage correct within achievement-level intervals		
Overall percentage correct †	*Basic* 146–177*	*Proficient* 178–209*	*Advanced* 210 and above*
41	43	60	75

†Includes twelfth-grade students who were below the *Basic* level.
*NAEP Science composite scale range.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Grade 12 Sample Question 8:

One characteristic that can be used to identify pure metals is density. If you determine the density of a pure metal, you can determine what the metal is, as shown in the table below.

Metal	Gold	Lead	Silver	Copper	Tin
Density (gram/cm^3)	19.3	11.3	10.5	8.9	7.3

Suppose that you have been given a ring and want to determine if it is made of pure gold. Design a procedure for determining the density of the ring. Explain the steps you would follow, including the equipment that you would use, and how you would use this equipment to determine the ring's density.

Responses to this question were scored according to a four-level rubric as Unsatisfactory/Incorrect, Partial, Essential, or Complete.

Sample "Complete" Response:

I would ~~a~~ determine the objects mass by using a scale. Then I would drop the object into a beaker of water and measure its displacement, which is its volume. I would the divide the mass by the volume

Sample "Essential" Response:

I would weigh it on a scale in grams. I would also place it in a beaker filled with water and see the displacement when it is added compared to the other rings.

Table 1.8 Sample Question 8 Results (Extended Constructed-Response)

Overall percentage "Essential" or better and percentages "Essential" or better within each achievement-level range: 2000

Grade 12	Percentage "Essential" or better within achievement-level intervals		
Overall percentage "Essential" or better †	*Basic* 146–177*	*Proficient* 178–209*	*Advanced* 210 and above*
19	18	58	89

† Includes twelfth-grade students who were below the *Basic* level.
*NAEP Science composite scale range.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Grade 12 Sample Question 9:

Based on your knowledge of genetics and the information in the preceding passage, answer the following questions.

What is a gene? What is it made of? What is the major function of a gene?

Responses to this question were scored according to a four-level rubric as Unsatisfactory/Incorrect, Partial, Essential, or Complete.

Sample "Complete" Response:

What is a gene? What is it made of? What is the major function of a gene?

a gene is a portion of DNA that codes for a particular trait or characteristic. a Gene is made of Deoxyribose Nucleic Acid. The major function of a gene is to code for particular protiens that are to be produced in a cell.

Sample "Essential" Response:

What is a gene? What is it made of? What is the major function of a gene?

Genes determine our characteristics and our traits. They are made up of strands of DNA.

Table 1.9 Sample Question 9 Results (Extended Constructed-Response)

Overall percentage "Essential" or better and percentages "Essential" or better within each achievement-level range: 2000

Grade 12	Percentage "Essential" or better within achievement-level intervals		
Overall percentage "Essential" or better †	*Basic* 146–177*	*Proficient* 178–209*	*Advanced* 210 and above*
22	24	44	56

†Includes twelfth-grade students who were below the *Basic* level.
*NAEP Science composite scale range.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Maps of Selected Item Descriptions

The science performance of fourth-, eighth-, and twelfth-graders can be illustrated by maps that position questions from the assessment onto the 0-to-300 scale. The resulting item maps are visual representations of how the difficulty of each question compares with a student's performance on the entire test.[21] The descriptions used on these maps focus on the science knowledge or skill needed to answer the question. For multiple-choice questions, the description indicates the knowledge or skill demonstrated by selection of the correct option; for constructed-response questions, the description takes into account the knowledge or skill specified by the different levels of scoring criteria for that question. Seven of the questions described on the item maps are included among the sample questions in the preceding section. Each of these sample questions is identified as such on the item map.

Figures 1.7 through 1.9 are item maps for grades 4, 8, and 12, respectively. For each question indicated on the map, students who scored above the scale point had a higher probability of successfully answering the question, and students who scored below the scale point had a lower probability of successfully answering the questions. The map location for each question identifies where that question was answered successfully by at least 65 percent of students for constructed-response questions and at least 74 percent of students for a four-option multiple-choice question.

As an example of how to interpret the item maps, consider the multiple-choice question in figure 1.7 that maps at score point 188. Fourth-graders were required to identify the function of a labeled human organ. Students who scored at or above 188 on the NAEP scale had a 74 percent chance of answering this question correctly. Students who scored below 188 had less than a 74 percent chance of doing so. This does not mean that all students scoring 188 or above always answered the question correctly, or that students scoring below 188 always answered the question incorrectly. Rather, the item map indicates a higher or lower probability of answering the question successfully depending on students' overall science knowledge and skills as measured by the NAEP scale.

[21] Details on the procedures used to develop item maps are provided in appendix A, pp. 174–175.

As another example of how to interpret the item maps, consider the question in figure 1.8 that maps at score point 194. Eighth-graders were asked to classify eight different organisms into two groups based on a physical characteristic. They were also asked to name a second physical characteristic they could have used. Students' responses to this constructed-response question were rated according to a four-level scoring guide that distinguished between "Unsatisfactory," "Partial," "Essential," and "Complete." As with all constructed-response questions portrayed on the item maps, the description of this item takes into account the requirements for a response to be rated at a certain level according to the scoring criteria for that question. With this question, the description is based on the level of performance required for a score of "Essential" or better. Students who scored at or above 194 on the NAEP scale had at least a 65 percent chance of demonstrating the knowledge and skill required to receive a rating of "Essential" or better on this question. Students who scored below 194 had less than a 65 percent chance of doing so.

Figure 1.7

Grade 4 Item Map

Map of selected item descriptions on the National Assessment of Educational Progress science scale for grade 4

This map describes the knowledge or skills associated with answering individual science questions. The map identifies the score point at which students had a high probability of successfully answering the question.*

300

260

256 Predict and explain the effect of fertilizer on algae growth in pond

250 *Recognize source of stored energy in beans*

242 Draw part of the food web of an ecosystem

235 Predict and explain cause of temperature increase in buckets of water left in sunlight

227 State and explain features of houses that combat heat or cold from weather

226 Explain roughly how to determine which of three bottles holds the most water—**Sample Question 3**

***Advanced* 205**

195 Recognize likelihood of rain based on type of cloud

188 *Identify function of labeled human organ*—**Sample Question 1**

185 Predict effect of one magnet on another

182 *Identify which animals produce very large number of offspring*

176 Recognize safest place to go in case one hears thunder and explain why

175 *Explain how bicycle reflectors work*

***Proficient* 170**

167 Explain the good and bad points of some given methods of garbage disposal

162 *Identify producer in given ecosystem*

160 *Identify forms of precipitation*

160 *Recognize cause of window rattling due to thunder*

151 *Recognize vibrations being common to several actions*

140 Name and explain some of the survival features of deer's body

***Basic* 138**

136 *Recognize best information for identifying a rock*

101 *Recognize instrument used to see stars*

93 Name one of the predators in given ecosystem

0

NOTE: Regular type denotes a constructed-response question. Italic type denotes a multiple-choice question.

* Each grade 4 science question in the 2000 assessment was mapped onto the NAEP 0–300 science scale. The position of the question on the scale represents the 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. Only selected questions from among those that were released after the 2000 assessment are presented. Scale score ranges for science achievement levels are referenced on the map.

To interpret the item map, consider, for example, the multiple-choice question that maps at a scale score of 250 for grade 4. This question concerns the source of stored energy in beans. Mapping the question at the 250 scale score indicates that at least 74 percent of the students performing at this point answered the question correctly. Among students with lower scores, less than 74 percent answered this question correctly.

SOURCE: National Center for Education Statistics. National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

NAEP Science Scale

Figure 1.8

Grade 8 Item Map

Map of selected item descriptions on the National Assessment of Educational Progress science scale for grade 8

This map describes the knowledge or skills associated with answering individual science questions. The map identifies the score point at which students had a high probability of successfully answering the question.*

300

260

250
251 Explain cause of echo in auditorium

240
241 Describe part of test to investigate behavior of paramecia
237 *Recognize reason for controlling air pressure in airplane*
235 State that length of shadow at noon is different in summer and winter

230
233 Identify top carnivore in pond ecosystem

220

Advanced **208**

210

. . . .

204 *Recognize tools useful for separating sand and salt mixture*—**Sample Question 4**

200
203 *Recognize appliance that converts energy to mechanical work*
197 *Recognize location of Moon with respect to Sun and Earth*

190
194 Classify living organisms based on physical characteristic—**Sample Question 6**
187 *Identify direction of river flow on contour map*

180
184 *Identify place in human body where digestion of protein occurs*

Proficient **170**

170

. . . .

166 Explain cause or effect of competition in pond ecosystem

160
160 Describe either production mechanism or method of travel of sound
156 *Identify why water is most important to living organisms*
155 *Identify effect of acid rain on pond ecosystem*

150
155 Explain that length of shadow relates to position of Sun in the sky

Basic **143**

. . . .

140
144 Explain reason or prevention methods of monument crumbling—**Sample Question 5**
137 Explain part of procedure of measuring change in length of shadow throughout the day

130
133 Discuss a factor to be considered in deciding to use ethyl alcohol as automobile fuel

120
121 Name some things that animals get from food for survival
117 Give partial explanation of how color of clothing relates to coolness on sunny day

110
111 *Identify organisms most likely to live in tropical rain forest*

100

0

NOTE: Regular type denotes a constructed-response question. Italic type denotes a multiple-choice question.

* Each grade 8 science question in the 2000 assessment was mapped onto the NAEP 0–300 science scale. The position of the question on the scale represents the 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. Only selected questions from among those that were released after the 2000 assessment are presented. Scale score ranges for science achievement levels are referenced on the map.

To interpret the item map, consider, for example, the constructed-response question that maps at a scale score of 194 at grade 8. This question concerns the classification of living organisms. Scoring of responses to this question allowed for partial credit by using a four-level scoring guide. Mapping the question at the 194 scale score indicates that at least 65 percent of the students performing at this point achieved a score of 2 ("Partial") on the question. Among students with lower scores, less than 65 percent received a score of 2 on the question.

SOURCE: National Center for Education Statistics. National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

NAEP Science Scale

Figure 1.9

Grade 12 Item Map

Map of selected item descriptions on the National Assessment of Educational Progress science scale for grade 12

This map describes the knowledge or skills associated with answering individual science questions. The map identifies the score point at which students had a high probability of successfully answering the question.*

300

260

250

252 *Identify relationships from given evolutionary tree*

240

230

232 *Predict volume of O_2 given parameters*

220

223 *Explain why Sun appears larger in January than in July*—**Sample Question 7**

219 *Name plants present when first amphibians appeared*

213 *Predict length of year on Mars*

212 *Estimate age of wood using radioactive dating*

210

***Advanced* 210**

200

199 *Explain cause of convection in atmosphere*

194 Make determination about composition of ring based on its density—**Sample Question 8**

190

187 Discuss cause or effect of genetic mutation

186 Name a disadvantage of using recombinant DNA technology

185 Explain how mountain forms near continental plate boundary

181 *Identify source of energy released in nuclear decay*

180

***Proficient* 178**

170

169 Describe similarities and differences between models of atom and solar system

166 *Name difference between offspring of sexually and asexually reproducing animals*

163 Draw simplified model of solar system

160

150

153 Predict distance of new planet from given and derived information

***Basic* 146**

140

130

126 Plot period v. distance from Sun for planets given data table

122 Name a way that the real solar system is different from the model drawn by the student

121 State cause and effect of AIDS

120

110

110 *Identify planet with longest year from data table*

100

90

80

80 Describe some modes of AIDS transmission

0

NOTE: Regular type denotes a constructed-response question. Italic type denotes a multiple-choice question.

* Each grade 12 science question in the 2000 assessment was mapped onto the NAEP 0–300 science scale. The position of the question on the scale represents the 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. Only selected questions from among those that were released after the 2000 assessment are presented. Scale score ranges for science achievement levels are referenced on the map.

To interpret the item map, consider, for example, the constructed-response question that maps at a scale score of 163 at grade 12. This question asked students to draw a simplified model of the solar system. Scoring of responses to this question allows for partial credit by using a three-level scoring guide. Mapping the question at the 163 scale score indicates that at least 65 percent of the students performing at this point achieved a score of 3 (“Complete”) on the question. Among students with lower scores, less than 65 percent received a score of 3 on the question.

SOURCE: National Center for Education Statistics. National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Interpreting NAEP Results

The average scores and percentages presented in this report are based on representative samples of students rather than on the entire population of students. Moreover, the collection of questions used at each grade level is but a sample of the many questions that could have been asked that measure the content and skills outlined in the NAEP science framework. As such, the results are subject to a measure of uncertainty, reflected in the standard error of the estimates. The standard errors for the estimated scale scores and percentages in this report are provided in appendix B.

The differences between scale scores and between percentages discussed in the following chapters take into account the standard errors associated with the estimates. Comparisons are based on statistical tests that consider both the magnitude of the difference between the group average scores or percentages and the standard errors of those statistics. Throughout this report, differences between scores and between percentages are pointed out only when they are significant from a statistical perspective. All differences reported are significant at the 0.05 level with appropriate adjustments for multiple comparisons. The term "significant" is not intended to imply a judgment about the absolute magnitude of the educational relevance of the differences. It is intended to identify statistically dependable population differences to help inform dialogue among policymakers, educators, and the public.

Readers are cautioned against interpreting NAEP results in a causal sense. Inferences related to subgroup performance or to the effectiveness of public and nonpublic schools, for example, should take into consideration the many socioeconomic and educational factors that may also impact on science performance.

Overview of Remaining Chapters

The results in chapters 2 and 3 of the report are based on a set of data with no accommodations offered to students. Findings are presented for the nation, for regions, for participating jurisdictions, and for the major reporting subgroups included in all NAEP report cards. Changes since the 1996 assessment are noted where the data permit comparisons. State-by-state results are included for the states and jurisdictions that participated in the science assessment at grades 4 and 8. Chapter 4 presents an overview of the second set of results—those that include students who were provided accommodations during the test administration. By including these results in the nation's science report card, the NAEP program continues a phased transition toward a more inclusive reporting sample. Future assessment results will be based solely on a student and school sample in which accommodations are permitted.

Chapter 5, which is based on the data with no accommodations offered, looks at factors that may influence teaching and learning, such as teacher certification and classroom practices. It includes information on the types of science courses students were taking at the time of the assessment.

This report also contains appendices that support or augment the results presented. Appendix A contains an overview of the NAEP science framework and specifications, information on the national and state samples, and a more detailed description of the major reporting subgroups featured in chapters 2 and 3. Appendix B contains the full data with standard errors for all tables and figures in this report. Appendix C presents selected state-level contextual variables from non-NAEP sources that may be associated with student performance. Appendix D contains a list of the NAEP science committee members.

Detailed information about the measurement methodology and data analysis techniques is available in the *NAEP 2000 Technical Report.*

2 Average Scale Score and Achievement-Level Results for the Nation and States

Chapter Focus

Are the nation's and states' fourth-, eighth-, and twelfth-graders making progress in science?

Chapter Contents

The extent to which the nation is realizing one of the goals set at the National Education Summit in 1989—to ensure that students leaving the fourth, eighth, and twelfth grades demonstrate competency in core subjects—can now be examined in light of results obtained from two administrations of the National Assessment of Educational Progress (NAEP) science assessment.[1] The results of the science assessment administered in 1996 showed 29 percent of fourth- and eighth-graders, and 21 percent of twelfth-graders demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.[2] Given the extensive push within the United States in the past decade to reform science teaching and learning, there is an interest in determining if the results of the NAEP 2000 science assessment, compared to the results from 1996, would positively reflect these reforms by showing an increase in the percentage of students demonstrating competency over challenging material.

This chapter presents the NAEP 2000 science results for the nation at grades 4, 8, and 12 and for participating states and jurisdictions at grades 4 and 8. Student performance on

---

1 U.S. Department of Education. (1991). *America 2000: An education strategy.* Washington, DC: Author.

2 Bourque, M. L., Champagne, A. B., & Crissman, S. (1997). *1996 science performance standards: Achievement results for the nation and the states.* Washington, DC: National Assessment Governing Board.

NAEP is reported in two ways: one, as average scores on the NAEP science composite scale, which ranges from 0 to 300, and two, as percentages of students who attained each of the three science achievement levels: *Basic, Proficient,* and *Advanced*. Discussion of students' progress over time is based on a comparison of the results in 2000 to those from the 1996 assessment. This comparison is possible because the assessments shared a common set of science questions based on the current science framework and because the populations of students were sampled and assessed using comparable procedures.

Readers are reminded that differences between scale scores and percentages discussed in this chapter take into account the standard errors associated with the estimates. Thus, a small difference between scores in one comparison may be significant while a similar or larger difference between scores in another comparison may not be statistically significant.

The results presented in this chapter are based on a representative sample of students assessed under conditions that did not permit accommodations for special-needs students. These were the same conditions under which the 1996 science assessment was administered, thus making it possible to report changes in student performance across the assessment years. A second set of results that reflect part of a phased transition toward a more inclusive reporting sample in which accommodations were permitted is presented in chapter 4.

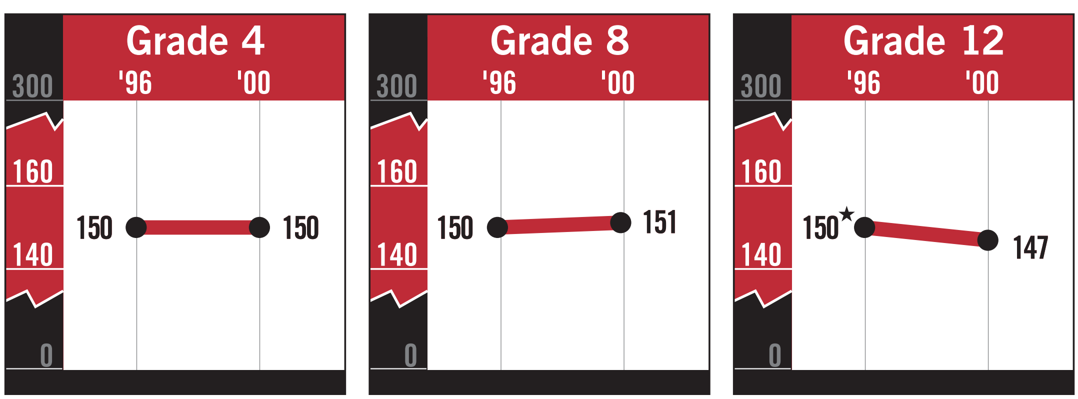
National Scale Score Results

Figure 2.1 presents the average science scale scores of fourth-, eighth-, and twelfth-grade students attending both public and nonpublic schools in 1996 and 2000. There were no statistically significant differences observed in average science scores from 1996 to 2000 at grades 4 and 8, and a decrease at grade 12 from an average score of 150 to 147.

Figure 2.1

National Scale Score Results

National average science scale scores, grades 4, 8, and 12: 1996 and 2000



★ Significantly different from 2000.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 and 2000 Science Assessments.

National Percentile Score Results

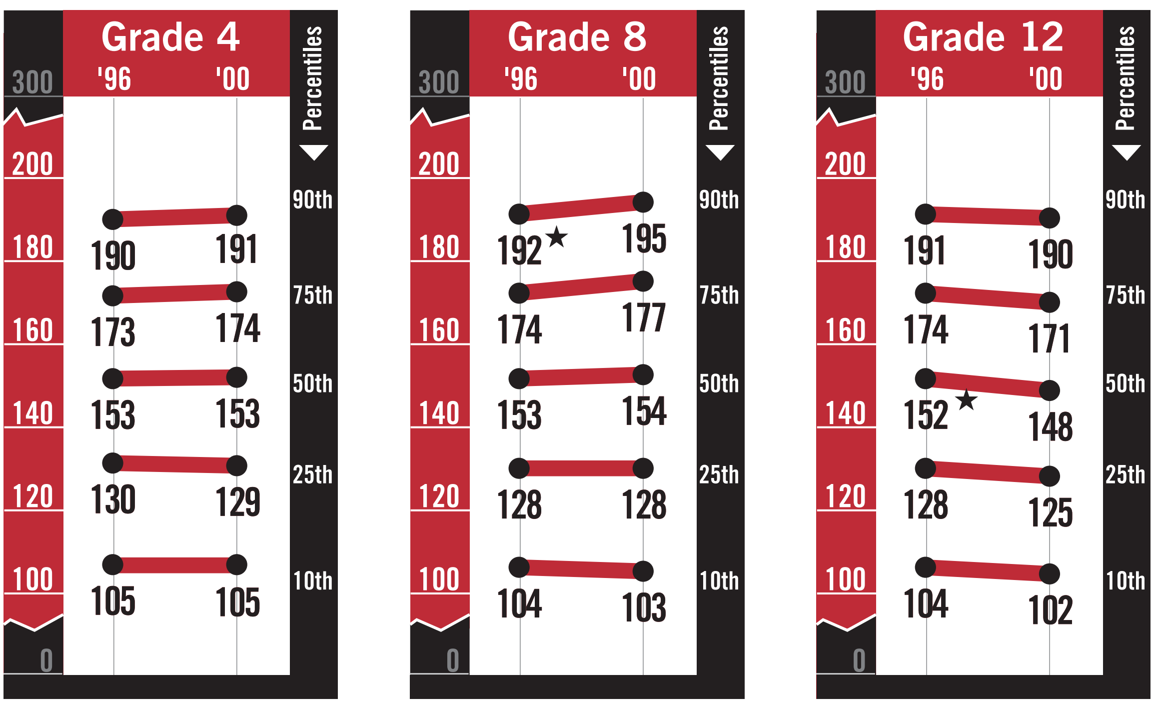
Changes in student performance can also be examined by looking at the percentile scores on the NAEP science scale across assessment years. The advantage of looking at data in this way is that it shows whether changes in the national average scores are reflected in scores across the performance distribution.

Figure 2.2 shows the science scores for grades 4, 8, and 12 at the 10th, 25th, 50th, 75th, and 90th percentiles in both 1996 and 2000. At grade 4, there was no significant difference observed in the percentile scores since 1996. Although there was no significant difference observed in the national average score at grade 8 between 1996 and 2000, there was an increase in the scale score at the 90th percentile—from 192 in 1996 to 195 in 2000—indicating improvement for the highest-performing students. At grade 12, the score at the 50th percentile declined between 1996 and 2000, indicating that the recent performance decline was primarily focused in the middle of the score distribution.

Figure 2.2

National Performance Distribution

National science scale score percentiles, grades 4, 8, and 12: 1996 and 2000



★ Significantly different from 2000.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 and 2000 Science Assessments.

Achievement-Level Results for the Nation

The science achievement levels—*Basic, Proficient,* and *Advanced*—used to report NAEP results were established by the National Assessment Governing Board (NAGB) in 1996. A discussion of the achievement-setting process can be found in chapter 1 of this report together with descriptions of what students in grades 4, 8, and 12 know and can do at each of the three achievement levels.

Achievement-level results for the nation's fourth-, eighth-, and twelfth-grade students are presented in figure 2.3. Results are presented in two ways: as the percentage of students within each achievement level interval, and as the percentage of students at or above the *Basic* and *Proficient* levels. It is necessary to keep in mind that the percentages at or above specific achievement levels are cumulative. Therefore, included among the percentage of students at or above the *Basic* level are also those who have achieved the *Proficient* and *Advanced* levels of performance, and included among students at or above the *Proficient* level are also those who have attained the *Advanced* level of performance.

As shown in figure 2.3, performance at or above the *Proficient* level—the achievement level identified by NAGB as the level that all students should reach—was attained by 29 percent of fourth-graders, 32 percent of eighth-graders, and 18 percent of twelfth-graders in 2000.

No statistically significant differences were detected on the NAEP measure at grade 4 (29 percent were at or above the *Proficient* level in both 1996 and 2000). However, at the eighth-grade level, some progress as demonstrated on the NAEP measure has been made. More students demonstrated competency over challenging science material; 32 percent were at or above the *Proficient* level in 2000 compared to 29 percent in 1996. At grade 12, the percentage of students at or above *Basic* declined between 1996 and 2000, from 57 percent to 53 percent. The apparent decline in the percentage of twelfth-graders at or above the *Proficient* level was not found to be statistically significant.

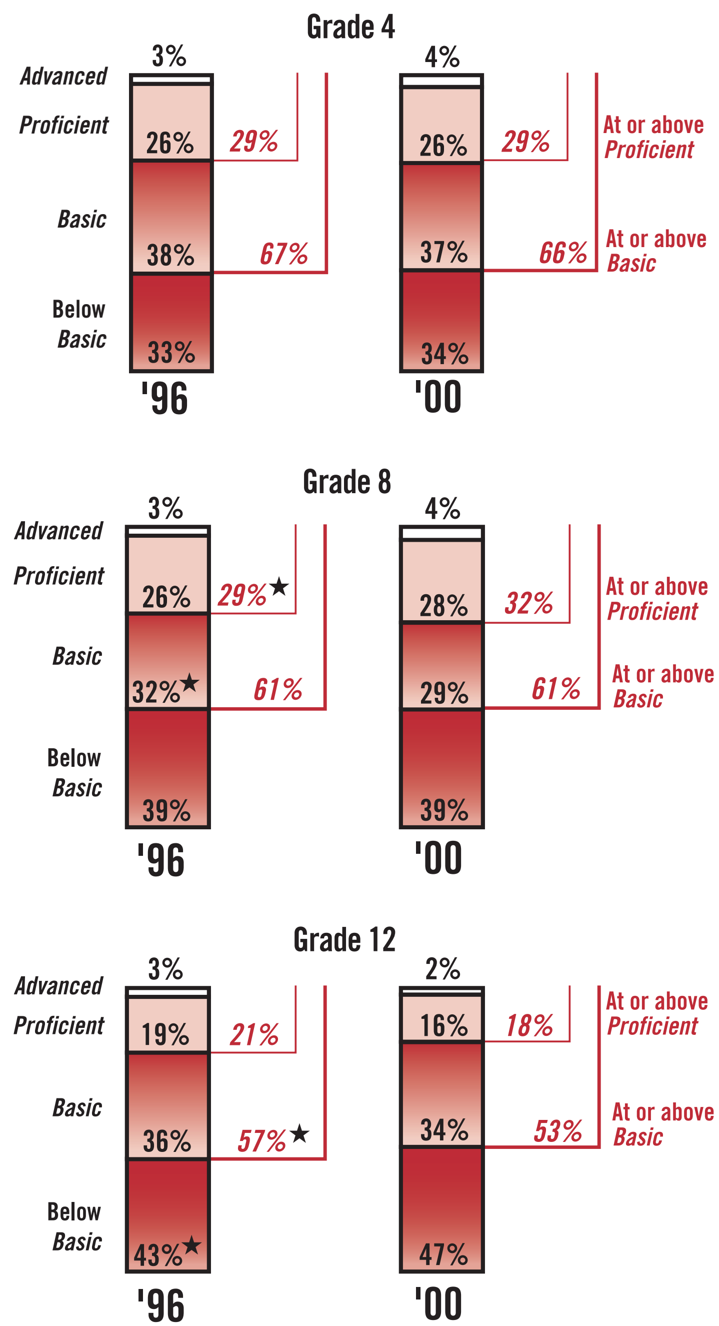
Figure 2.3

National Achievement-Level Results

Percentage of students within each science achievement-level range and at or above achievement levels, grades 4, 8, and 12: 1996 and 2000

How to read these figures:

- The italicized percentages to the right of the shaded bars represent the percentages of students at or above *Basic* and *Proficient*.
- The percentages in the shaded bars represent the percentages of students within each achievement level.



★ Significantly different from 2000.

NOTE: Percentages within each science achievement level range may not add to 100, or to the exact percentages at or above achievement levels, due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 and 2000 Science Assessments.

Results for Regions of the Country

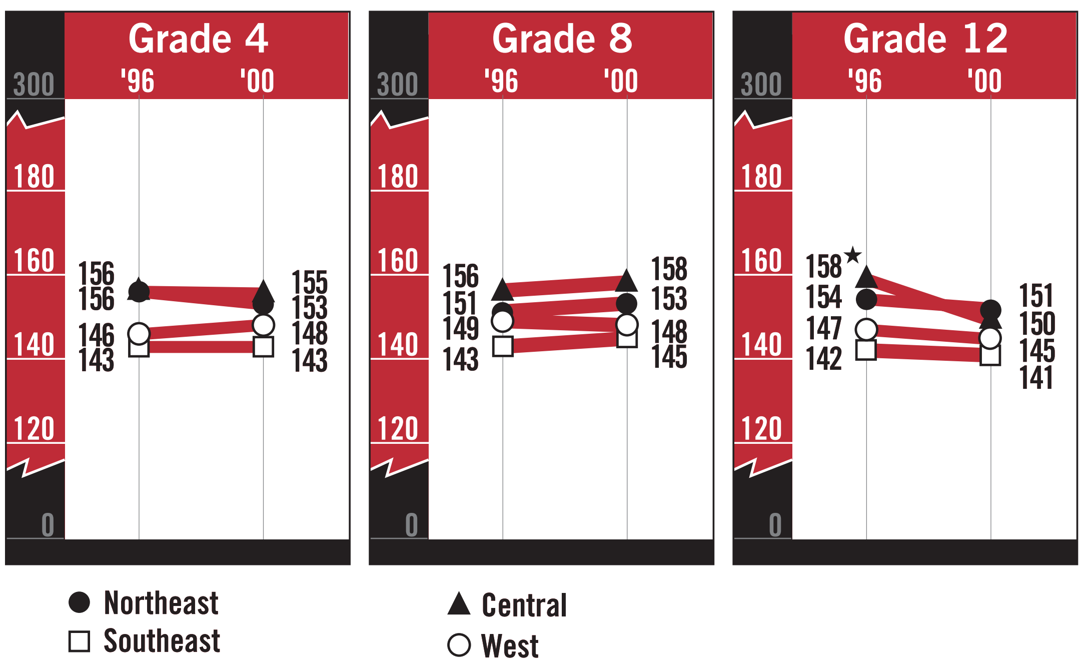
This section examines results for four regions of the country: Northeast, Southeast, Central, and West. A listing of the states and other jurisdictions within these regions is provided in appendix A. Figure 2.4 presents scale score results by region. At grades 4 and 8, there were no statistically significant differences in the performance of students attending schools in the Northeast, Southeast, Central, and West regions between 1996 and 2000. At grade 12, however, the average science score for students attending schools in the Central region was lower in 2000 than in 1996.

Comparisons between the regions in the 2000 assessment show that fourth-grade students attending schools in the Northeast and Central regions outperformed their peers in the West and Southeast. In addition, grade 4 students in the West had higher scores than students in the Southeast. Eighth-grade students attending schools in the Northeast and Central regions had higher average scores than their peers in the West and Southeast. Twelfth-grade students attending schools in the Southeast were outperformed by their peers in the Northeast and Central regions.

Figure 2.4

National Scale Score Results by Region

National science scale score results by region of the country, grades 4, 8, and 12: 1996 and 2000



★ Significantly different from 2000.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 and 2000 Science Assessments.

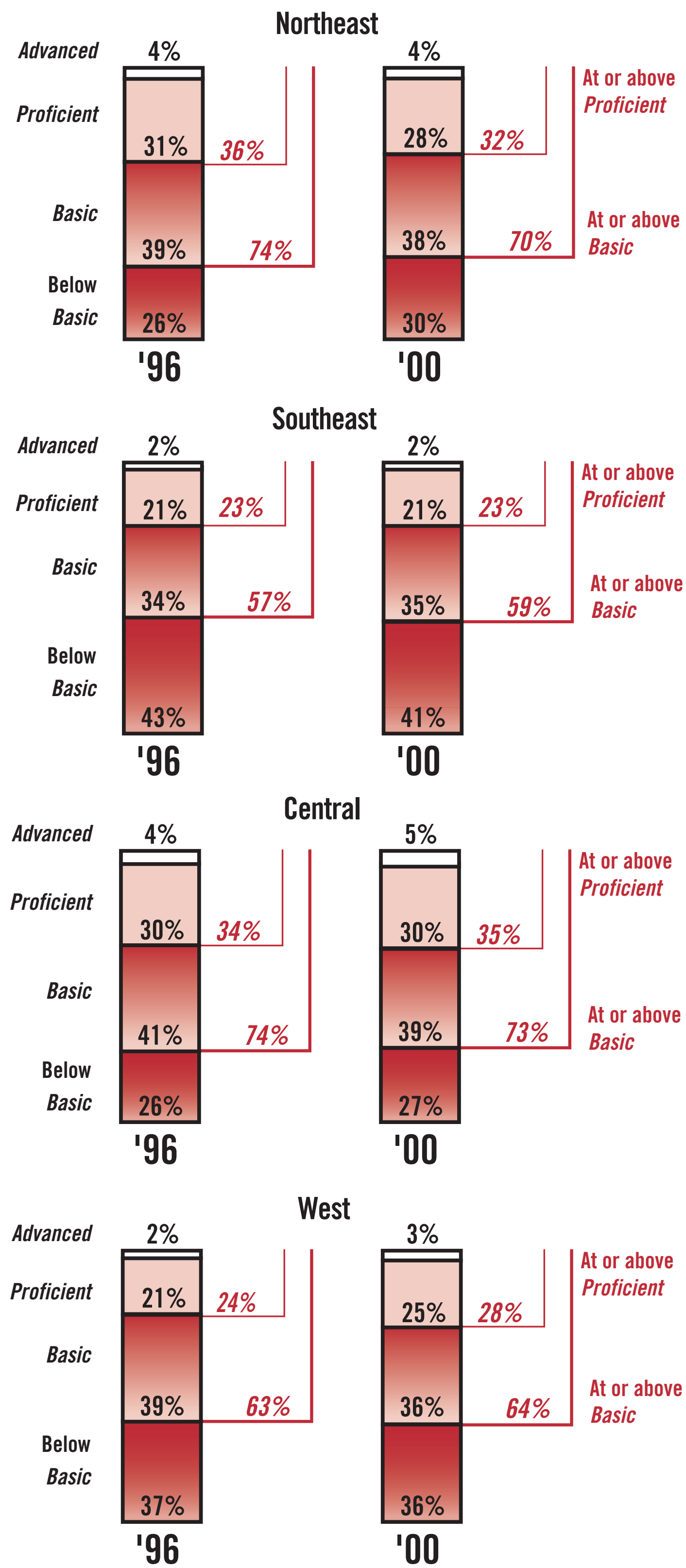
Figure 2.5 presents the achievement-level results by region. At grades 4 and 8, there were no statistically significant changes in the percentages of students at or above the *Basic* and *Proficient* levels between 1996 and 2000 in any of the four regions. The one percentage point increase at the *Advanced* level in the Southeast at grade 8 was, however, statistically significant. At grade 12, the percentage of students at or above the *Basic* and *Proficient* levels decreased in the Central region between 1996 and 2000.

A number of differences can be seen when the results for each of the three grades in 2000 are compared between the regions. At grade 4, both the Northeast and Central regions had higher percentages of students at or above the *Basic* level than in the Southeast and West, and higher percentages at or above *Proficient* than in the Southeast. At grade 8, both the Northeast and the Central regions had higher percentages of students at or above the *Basic* and *Proficient* levels than did the Southeast and West. In addition, the percentage of eighth-graders at or above the *Basic* level in the Central region was higher than the percentage of eighth-graders in the Northeast. At grade 12, the Northeast and Central regions had higher percentages of students at or above the *Basic* level than did the Southeast. There was no statistically significant difference between the regions in the percentage of students at or above the *Proficient* level at grade 12.

Figure 2.5a

National Achievement-Level Results by Region

Percentage of students within each science achievement-level range and at or above achievement levels, by region of the country, grade 4: 1996 and 2000

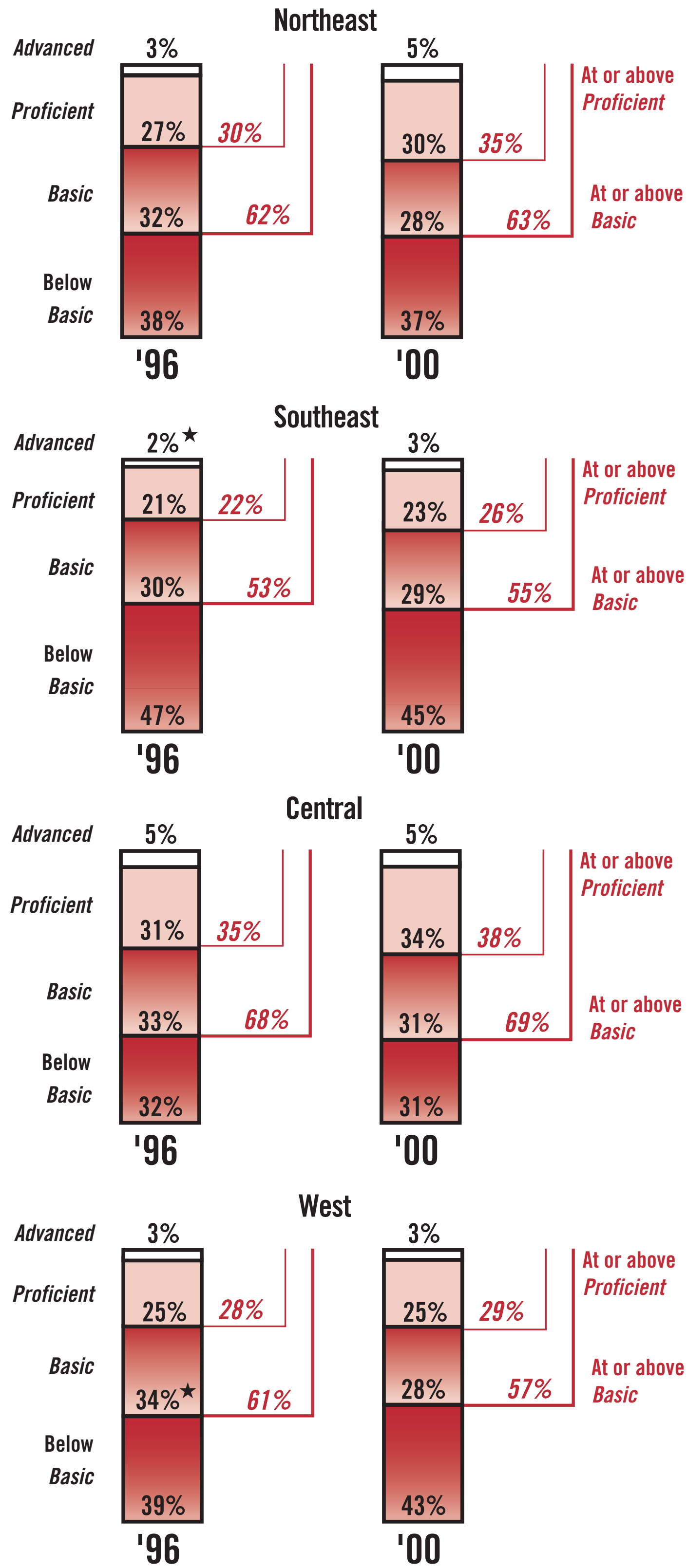


See footnotes at end of figure. ▶

Figure 2.5b

National Achievement-Level Results by Region (continued)

Percentage of students within each science achievement-level range and at or above achievement levels, by region of the country, grade 8: 1996 and 2000

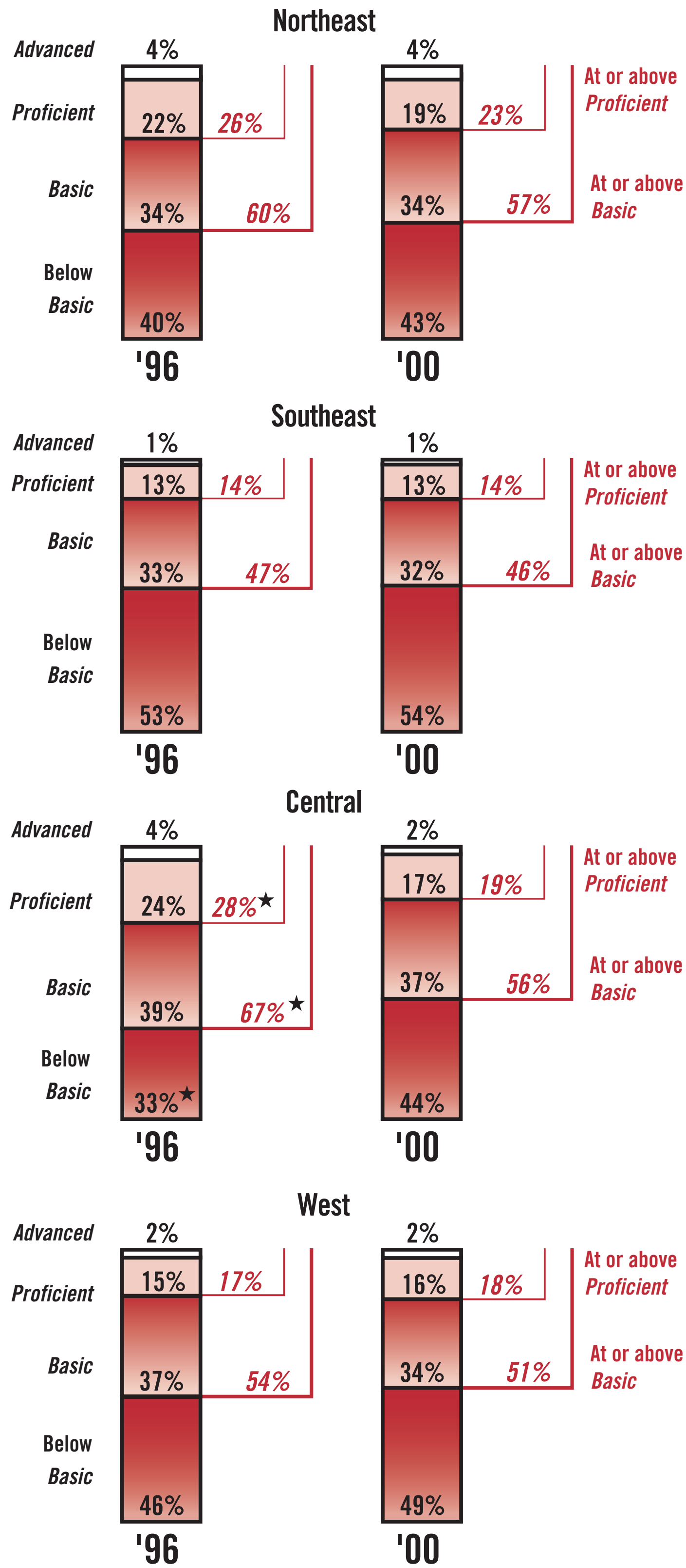


See footnotes at end of figure. ▶

Figure 2.5c

National Achievement-Level Results by Region (continued)

Percentage of students within each science achievement-level range and at or above achievement levels, by region of the country, grade 12: 1996 and 2000



★ Significantly different from 2000.

NOTE: Percentages within each science achievement level range may not add to 100, or to the exact percentages at or above achievement levels, due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 and 2000 Science Assessments.

State Results

In addition to the national results, science performance data were also collected for students in grades 4 and 8 who attended public schools in states and other jurisdictions that chose to participate in the assessment. Although 45 states and jurisdictions participated at grade 4, and 44 states and jurisdictions participated at grade 8, not all met minimum school participation guidelines for reporting their results in 2000. (See appendix A for details on participation and reporting guidelines.) Results from the 2000 assessment for grades 4 and 8 in Wisconsin and for grade 8 in the Virgin Islands are not included in this report because they failed to meet the minimum public school participation rate of 70 percent. Jurisdictions that failed to meet one or more of the other participation guidelines are noted in each of the tables. Results from both the 1996 and 2000 state assessments are presented for grade 8, but results from 2000 only are reported at grade 4 since there was no state-level assessment administered to fourth-graders in 1996. Tables presenting state-level results at grade 8, indicate statistically significant changes across years when examining only one jurisdiction at a time (⋆), and when using a multiple comparison procedure based on all the jurisdictions that participated (‡). Only those differences based on the multiple comparison procedure are discussed.

In examining the "accommodations-not-permitted" results for jurisdictions presented in this chapter, it should be noted that schools participating in the NAEP assessments under these conditions were permitted to exclude those students who could not be assessed meaningfully without accommodations. Exclusion rates vary across jurisdictions not only because of differences in the implementation of the Individuals with Disabilities Education Act (IDEA), but also because of population shifts in the percentage of students classified with disabilities (SD) and, especially, limited English proficient (LEP) students. Therefore, comparisons of assessment results across jurisdictions and within jurisdictions across years should be made with caution. The percentage of students excluded from the assessment has implications for the representativeness of the sample assessed within a jurisdiction. No adjustments have been made for differing exclusion rates across jurisdictions or across years. Thus, a comparison within a jurisdiction across years or between two jurisdictions may be based on samples with exclusion rates that differ considerably. The exclusion rates for each jurisdiction are presented in appendix A.

Scale Score Results by Jurisdiction

The average scale scores of public school students for participating jurisdictions are presented in table 2.1 for grade 4 and table 2.2 for grade 8. Whereas the national results shown in previous sections of this chapter represent both public and nonpublic schools combined, the national average scores shown in each of these tables represent the performance of public school students only. Of the 36 jurisdictions that participated in both the 1996 and 2000 state level assessments at grade 8, three showed score gains in 2000: Missouri and the Department of Defense domestic schools and overseas schools (DDESS and DoDDS).

Table 2.1 State Average Score Results, Grade 4

Average science scale score results by state for grade 4 public schools: 2000

Nation	148
Alabama	143
Arizona	141
Arkansas	144
California †	131
Connecticut	156
Georgia	143
Hawaii	136
Idaho †	153
Illinois †	151
Indiana †	155
Iowa †	160
Kentucky	152
Louisiana	139
Maine †	161
Maryland	146
Massachusetts	162
Michigan †	154
Minnesota †	157
Mississippi	133
Missouri	156
Montana †	160
Nebraska	150
Nevada	142
New Mexico	138
New York †	149
North Carolina	148
North Dakota	160
Ohio †	154
Oklahoma	152
Oregon †	150
Rhode Island	148
South Carolina	141
Tennessee	147
Texas	147
Utah	155
Vermont †	159
Virginia	156
West Virginia	150
Wyoming	158
Other Jurisdictions	
American Samoa	51
DDESS	157
DoDDS	156
Guam	110
Virgin Islands	116

† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation.
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools.
DoDDS: Department of Defense Dependents Schools (Overseas).
NOTE: National results are based on the national sample, not on aggregated state assessment samples.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Table 2.2 State Average Score Results, Grade 8

Average science scale score results by state for grade 8 public schools: 1996 and 2000

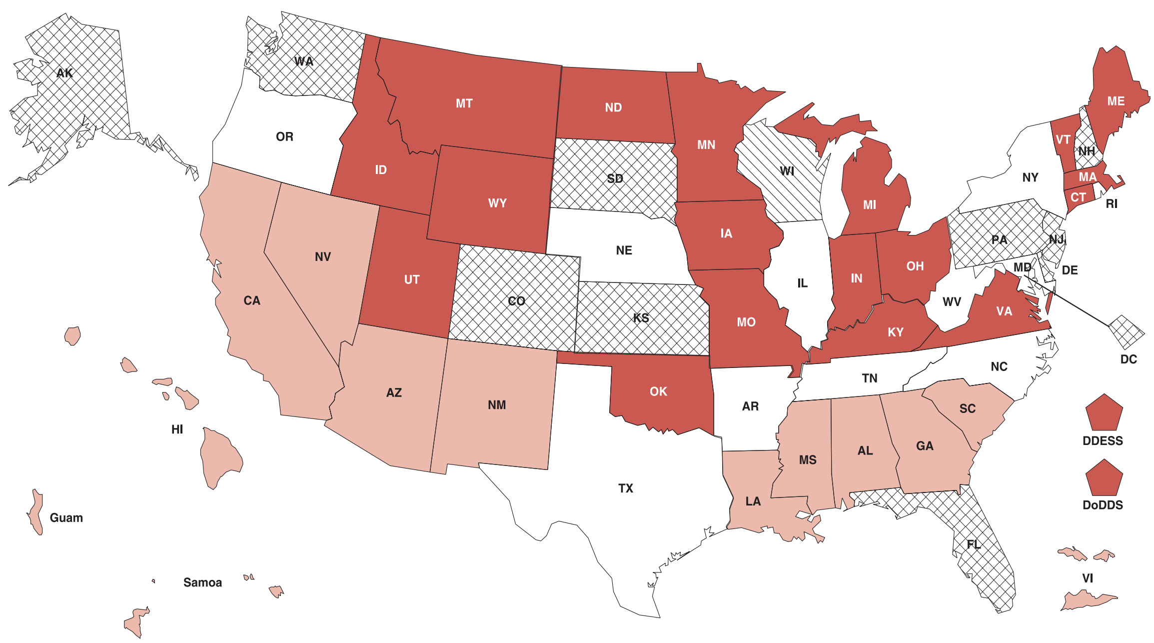
	1996	2000
Nation	148	149
Alabama	139	141
Arizona †	145	146
Arkansas	144	143
California †	138 *	132
Connecticut	155	154
Georgia	142	144
Hawaii	135	132
Idaho †	—	159
Illinois †	—	150
Indiana †	153	156
Kentucky	147 *	152
Louisiana	132	136
Maine †	163 *	160
Maryland	145	149
Massachusetts	157	161
Michigan †	153	156
Minnesota †	159	160
Mississippi	133	134
Missouri	151 ‡	156
Montana †	162	165
Nebraska	157	157
Nevada	—	143
New Mexico	141	140
New York †	146	149
North Carolina	147	147
North Dakota	162	161
Ohio	—	161
Oklahoma	—	149
Oregon †	155	154
Rhode Island	149	150
South Carolina	139	142
Tennessee	143	146
Texas	145	144
Utah	156	155
Vermont †	157 *	161
Virginia	149	152
West Virginia	147	150
Wyoming	158	158
Other Jurisdictions		
American Samoa	—	72
DDESS	153 ‡	159
DoDDS	155 ‡	159
Guam	120	114

* Significantly different from 2000 if only one jurisdiction or the nation is being examined.
‡ Significantly different from 2000 when examining only one jurisdiction and when using a multiple comparison procedure based on all jurisdictions that participated both years.
† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2000.
— Indicates that the jurisdiction did not participate.
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools. DoDDS: Department of Defense Dependents Schools (Overseas).
NOTE: National results are based on the national sample, not on aggregated state assessment samples.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 and 2000 Science Assessments.

The maps in figures 2.6 and 2.7 compare state and national average scores at grades 4 and 8, respectively. At grade 4, 20 jurisdictions had scores that were higher than the national average score, 13 had scores that were lower than the national average, and no statistically significant differences were detected between the state and national average for 11 states. At grade 8, 18 jurisdictions had scores that were higher than the national average score, 13 had scores that were lower than the national average, and no significant differences were detected between the state and national average for 11 states.

Figure 2.6
State v. National Scale Score, Grade 4

Comparison results of state and national average science scale scores for grade 4: 2000



- State has higher average scale score than nation.
- State average score was not found to be significantly different from the nation.
- State has lower average scale score than nation.
- State did not meet the minimum participation rate guidelines (see technical notes).
- State did not participate in the NAEP 2000 science state assessment.

NOTE: Caution should be exercised when interpreting comparisons among states and other jurisdictions. NAEP performance estimates are not adjusted to account for the socioeconomic or demographic, differences among states and jurisdictions.

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools.

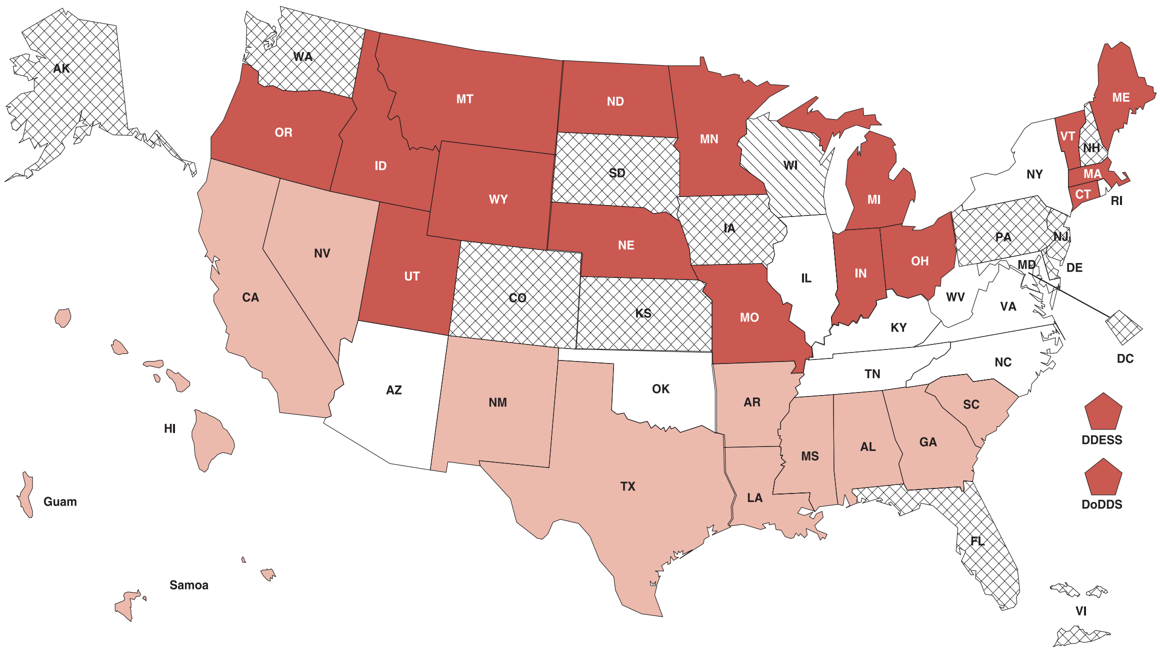
DoDDS: Department of Defense Dependents Schools (Overseas).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Figure 2.7

State v. National Scale Score, Grade 8

Comparison results of state and national average science scale scores for grade 8: 2000



- State has higher average scale score than nation.
- State average score was not found to be significantly different from the nation.
- State has lower average scale score than nation.
- State did not meet the minimum participation rate guidelines (see technical notes).
- State did not participate in the NAEP 2000 science state assessment.

NOTE: Caution should be exercised when interpreting comparisons among states and other jurisdictions. NAEP performance estimates are not adjusted to account for the socioeconomic or demographic, differences among states and jurisdictions.

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools.

DoDDS: Department of Defense Dependents Schools (Overseas).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

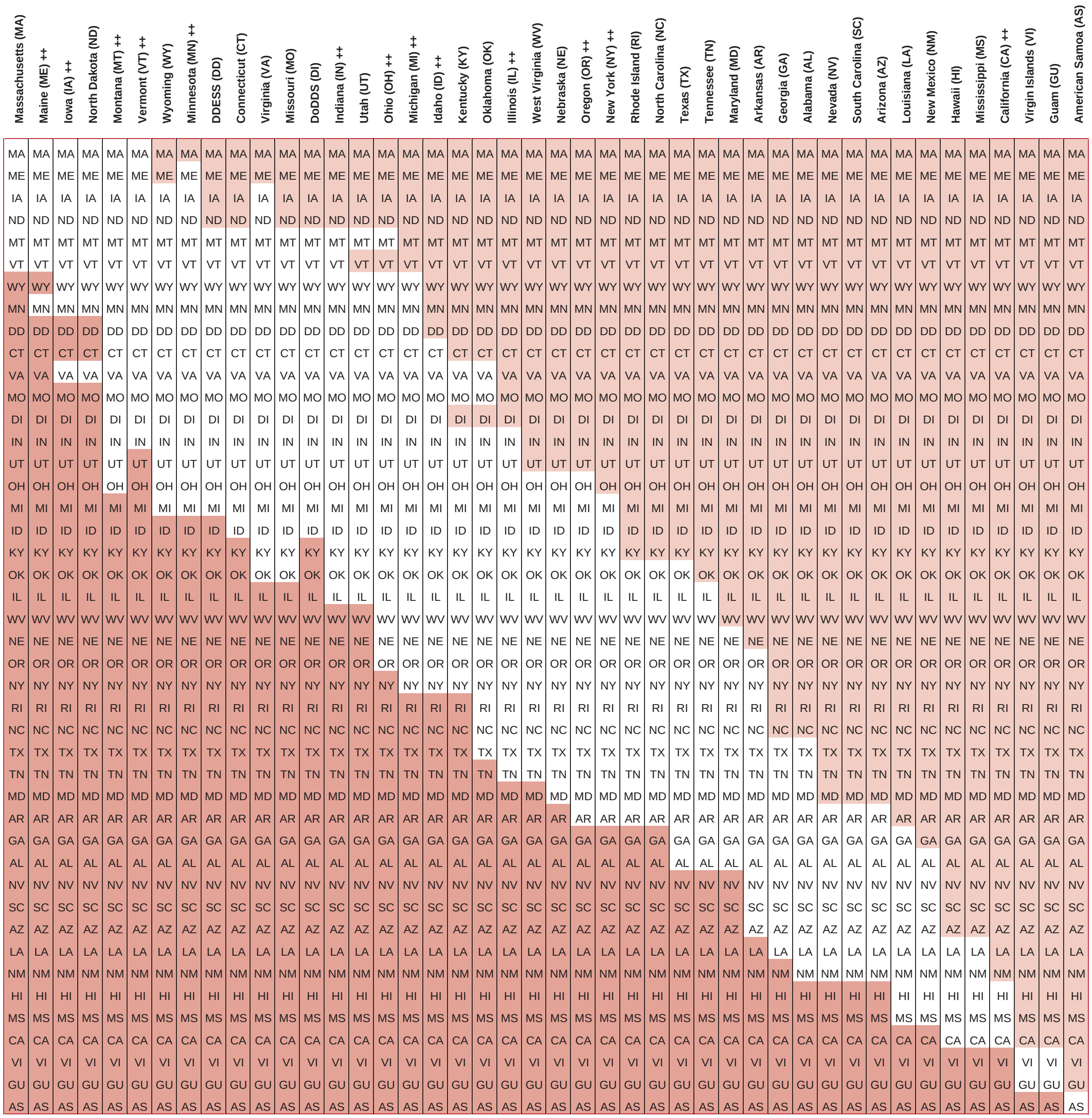
Cross-State Scale Score Comparisons

Figures 2.8 and 2.9 display the differences between the scale scores for all possible pairings of participating jurisdictions at grades 4 and 8, respectively. The variation in shading indicates whether a jurisdiction listed across the top of the figure had a score that was higher than, lower than, or not significantly different from other jurisdictions. Within each figure, jurisdictions are ranked from highest to lowest average scale score, both from left to right across the columns and down the rows. For example in figure 2.8, the first cell in the second row compares the average score at grade 4 in Massachusetts (MA) to the average score in Maine (ME). The lack of shading in this cell indicates that there was no statistically significant difference found between the scores in these two states. Moving down the first column to Wyoming (WY), the shading changes to indicate that the average score in Massachusetts was higher than that in Wyoming. At grade 4, the top 6 states had average scores that were not found to differ significantly from one another. These states were Iowa, Maine, Massachusetts, Montana, North Dakota, and Vermont. At grade 8, Idaho, Maine, Massachusetts, Minnesota, North Dakota, Ohio, Vermont, and the Department of Defense domestic schools and overseas schools all performed similarly (i.e., no significant differences were detected in the average scores of these 9 jurisdictions) and were only outperformed by Montana.

Figure 2.8: Cross-State Scale Score Comparisons, Grade 4

Comparisons of average science scale scores for grade 4 public schools: 2000

Instructions: Read down the column directly under a jurisdiction name listed in the heading at the top of the figure. Match the shading intensity surrounding a jurisdiction's abbreviation to the key below to determine whether the average science scale score of this jurisdiction is higher than, the same as, or lower than the jurisdiction in the column heading. For example, in the column under Wyoming, Wyoming's score was lower than Massachusetts and Maine, about the same as all the states from Iowa through Michigan, and higher than the remaining states down the column.



Jurisdiction has statistically significantly higher average scale score than the jurisdiction listed at the top of the figure.

No statistically significant difference detected from the jurisdiction listed at the top of the figure.

Jurisdiction has statistically significantly lower average scale score than the jurisdiction listed at the top of the figure.

The between jurisdiction comparisons take into account sampling and measurement error and that each jurisdiction is being compared with every other jurisdiction. Significance is determined by an application of a multiple comparison procedure (see appendix A).

++Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see appendix A).

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools.

DoDDS: Department of Defense Dependents Schools (Overseas).

NOTE: Differences between states and jurisdictions may be partially explained by other factors not included in this figure.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Figure 2.9: Cross-State Scale Score Comparisons, Grade 8

Comparisons of average science scale scores for grade 8 public schools: 2000

Instructions: Read down the column directly under a jurisdiction name listed in the heading at the top of the figure. Match the shading intensity surrounding a jurisdiction's abbreviation to the key below to determine whether the average science scale score of this jurisdiction is higher than, the same as, or lower than the jurisdiction in the column heading. For example, in the column under Indiana, Indiana's score was lower than Montana, Massachusetts, Vermont, and North Dakota, about the same as all the states from Ohio through Kentucky, and higher than the remaining states down the column.

Montana (MT) ++	Massachusetts (MA)	Vermont (VT) ++	North Dakota (ND)	Ohio (OH)	Minnesota (MN) ++	Maine (ME) ++	DoDDS (DI)	Idaho (ID) ++	DDESS (DD)	Wyoming (WY)	Nebraska (NE)	Missouri (MO)	Michigan (MI) ++	Indiana (IN) ++	Utah (UT)	Oregon (OR) ++	Connecticut (CT)	Virginia (VA)	Kentucky (KY)	Illinois (IL) ++	Rhode Island (RI)	West Virginia (WV)	Oklahoma (OK)	Maryland (MD)	New York (NY) ++	North Carolina (NC)	Tennessee (TN)	Arizona (AZ) ++	Texas (TX)	Georgia (GA)	Nevada (NV)	Arkansas (AR)	South Carolina (SC)	Alabama (AL)	New Mexico (NM)	Louisiana (LA)	Mississippi (MS)	Hawaii (HI)	California (CA) ++	Guam (GU)	American Samoa (AS)
MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT
MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA	MA
VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT
ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH	OH
MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN	MN
ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME	ME
DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI	DI
ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID
DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD
WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY	WY
NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO	MO
MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI	MI
IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN	IN
UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT	UT
OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR	OR
CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT
VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA	VA
KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY	KY
IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL	IL
RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI	RI
WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV	WV
OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD
NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY	NY
NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN	TN
AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ	AZ
TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX	TX
GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA	GA
NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV	NV
AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR	AR
SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC	SC
AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL	AL
NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM
LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA	LA
MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS	MS
HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI	HI
CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA
GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU	GU
AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS

Jurisdiction has statistically significantly higher average scale score than the jurisdiction listed at the top of the figure.

No statistically significant difference detected from the jurisdiction listed at the top of the figure.

Jurisdiction has statistically significantly lower average scale score than the jurisdiction listed at the top of the figure.

The between jurisdiction comparisons take into account sampling and measurement error and that each jurisdiction is being compared with every other jurisdiction. Significance is determined by an application of a multiple comparison procedure (see appendix A).

++Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see appendix A).

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools.

DoDDS: Department of Defense Dependents Schools (Overseas).

NOTE: Differences between states and jurisdictions may be partially explained by other factors not included in this figure.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Achievement-Level Results by Jurisdiction

Like the national results, achievement-level results for jurisdictions are presented in two ways: the percentage of students within each science achievement-level range, and the percentage of students at or above the *Proficient* level. The percentage of students within each science achievement-level range in 2000 by jurisdiction is presented in figure 2.10 for grade 4 and figure 2.11 for grade 8. The shaded bars represent the proportion of students in each of the three achievement levels (*Basic, Proficient,* and *Advanced)* as well as the proportion of students who are below *Basic.* Each population of students is aligned at the point where the *Proficient* level begins, so that scanning down the horizontal bars allows for easy comparison of the percentages of students who were at or above *Proficient*. Jurisdictions are listed in the figures in three clusters based on a statistical comparison of the percentage of students at or above *Proficient* in each jurisdiction with the national percentage of public school students who were at or above *Proficient*. The cluster of jurisdictions at the top of each figure had a higher percentage of students at or above the *Proficient* level compared to the nation. For jurisdictions in the middle cluster, the percentages of students did not differ significantly from the national percentage. Jurisdictions in the bottom cluster had percentages lower than the national percentage. Within each cluster, jurisdictions are listed in alphabetical order.

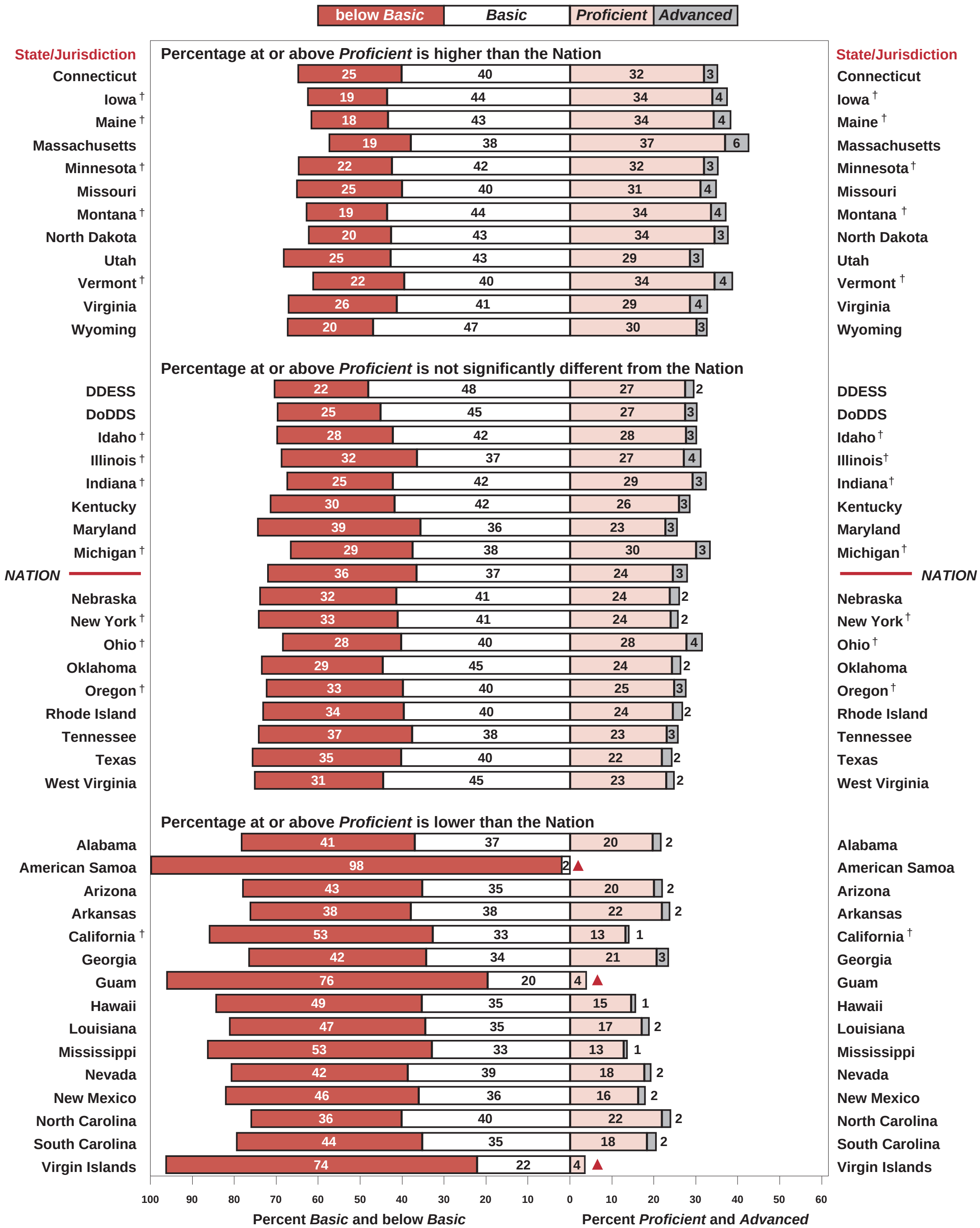
Figure 2.10 shows that at grade 4, 12 jurisdictions had higher percentages of students at or above *Proficient* than the nation, 17 had percentages that were not different from the nation, and 15 had percentages that were lower than the nation. In figure 2.11, the results for grade 8 show 17 jurisdictions with higher percentages of students at or above *Proficient* than the nation, 8 with percentages that were not different from the nation, and 17 with percentages that were lower than the nation.

Figure 2.10

State Achievement-Level Results, Grade 4

Percentage of students within each science achievement-level range by state for grade 4 public schools: 2000

The bars below indicate the percentages of students in each NAEP science achievement level. Each population of students is aligned at the point where the *Proficient* level begins, so that they may be compared at *Proficient* and above. States are listed alphabetically within three groups: the percentage at or above *Proficient* is higher than, not significantly different from, or lower than the nation.



† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation.

▲ Percentage is between 0.0 and 0.5.

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools. DoDDS: Department of Defense Dependents Schools (Overseas).

NOTE: Numbers may not add to 100 due to rounding. National results are based on the national sample, not on aggregated state assessment samples.

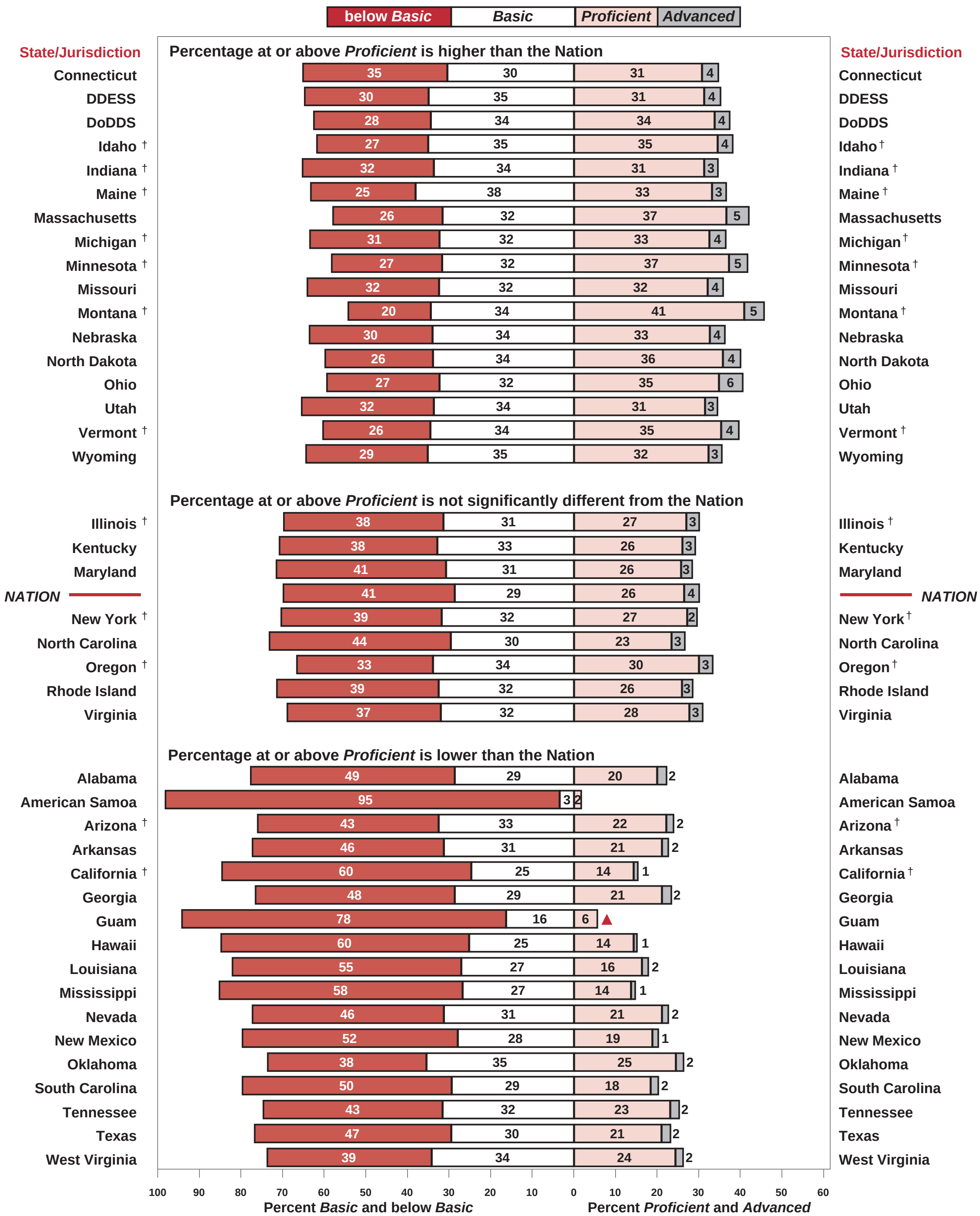
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Figure 2.11

State Achievement-Level Results, Grade 8

Percentage of students within each science achievement-level range by state for grade 8 public schools: 2000

The bars below indicate the percentages of students in each NAEP science achievement level. Each population of students is aligned at the point where the *Proficient* level begins, so that they may be compared at *Proficient* and above. States are listed alphabetically within three groups: the percentage at or above *Proficient* is higher than, not significantly different from, or lower than the nation.



† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation.
▲ Percentage is between 0.0 and 0.5.
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools. DoDDS: Department of Defense Dependents Schools (Overseas).
NOTE: Numbers may not add to 100 due to rounding. National results are based on the national sample, not on aggregated state assessment samples.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Tables 2.3 and 2.4 present the percentages of students performing at or above the *Proficient* level by jurisdiction for grades 4 and 8, respectively. At grade 4, the percentage of students at or above *Proficient* ranged from less than 1 percent to 43 percent of students in 2000. At grade 8, the percentage of students at or above the *Proficient* level ranged from 2 percent to 46 percent in 2000. Of the 36 jurisdictions that participated in both 1996 and 2000 at grade 8, 6 made gains in the percentage of students at or above *Proficient*: Kentucky, Missouri, Vermont, West Virginia, and the Department of Defense domestic schools and overseas schools (DDESS and DoDDS).

Table 2.3 State *Proficient* Level Results, Grade 4

Percentage of students at or above the *Proficient* level in science by state for grade 4 public schools: 2000

Nation	28
Alabama	22
Arizona	22
Arkansas	24
California †	14
Connecticut	35
Georgia	23
Hawaii	16
Idaho †	30
Illinois †	31
Indiana †	32
Iowa †	37
Kentucky	29
Louisiana	19
Maine †	38
Maryland	26
Massachusetts	43
Michigan †	33
Minnesota †	35
Mississippi	14
Missouri	35
Montana †	37
Nebraska	26
Nevada	19
New Mexico	18
New York †	26
North Carolina	24
North Dakota	38
Ohio †	31
Oklahoma	26
Oregon †	28
Rhode Island	27
South Carolina	21
Tennessee	26
Texas	24
Utah	32
Vermont †	39
Virginia	33
West Virginia	25
Wyoming	33
Other Jurisdictions	
American Samoa	▲
DDESS	29
DoDDS	30
Guam	4
Virgin Islands	4

† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation.
▲ Percentage is between 0.0 and 0.5.
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools.
DoDDS: Department of Defense Dependents Schools (Overseas).
NOTE: National results are based on the national sample and not on aggregated state assessment samples.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Table 2.4 State *Proficient* Level Results, Grade 8

Percentage of students at or above the *Proficient* level in science by state for grade 8 public schools: 1996 and 2000

	1996	2000
Nation	27	30
Alabama	18 *	22
Arizona †	23	24
Arkansas	22	23
California †	20	15
Connecticut	36	35
Georgia	21	23
Hawaii	15	15
Idaho †	—	38
Illinois †	—	30
Indiana †	30	35
Kentucky	23 ‡	29
Louisiana	13 *	18
Maine †	41	37
Maryland	25	28
Massachusetts	37 *	42
Michigan †	32	37
Minnesota †	37	42
Mississippi	12	15
Missouri	28 ‡	36
Montana †	41	46
Nebraska	35	36
Nevada	—	23
New Mexico	19	20
New York †	27	30
North Carolina	24	27
North Dakota	41	40
Ohio	—	41
Oklahoma	—	26
Oregon †	32	33
Rhode Island	26	29
South Carolina	17	20
Tennessee	22	25
Texas	23	23
Utah	32	34
Vermont †	34 ‡	40
Virginia	27	31
West Virginia	21 ‡	26
Wyoming	34	36
Other Jurisdictions		
American Samoa	—	2
DDESS	27 ‡	35
DoDDS	31 ‡	37
Guam	7	6

* Significantly different from 2000 if only one jurisdiction or the nation is being examined.
‡ Significantly different from 2000 when examining only one jurisdiction and when using a multiple comparison procedure based on all jurisdictions that participated both years.
† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation.
— Indicates that the jurisdiction did not participate.
NOTE: National results are based on the national sample and not on aggregated state assessment samples.
DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools.
DoDDS: Department of Defense Dependents Schools (Overseas).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 and 2000 Science Assessments.

Cross-State Achievement Level Comparisons

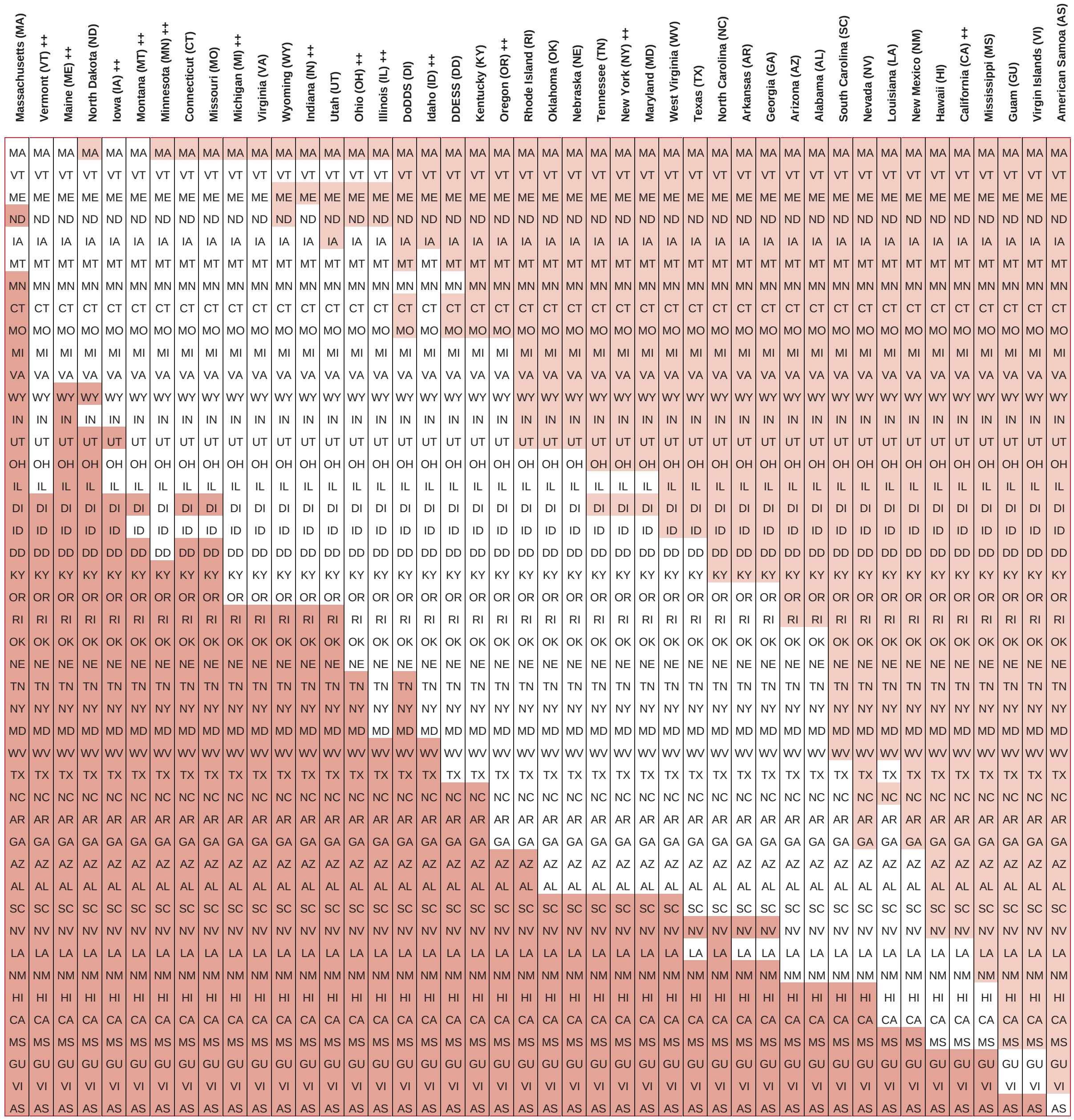
Figures 2.12 and 2.13 display the same type of state comparisons presented earlier for scale score results, but this time the performance measure being compared is the percentage of students at or above the *Proficient* achievement level for grades 4 and 8, respectively.

At grade 4, there were five states that had higher percentages of students at or above *Proficient* than the other states, but for which no significant differences were observed between them: Iowa, Maine, Massachusetts, Montana, and Vermont. At grade 8, the highest percentages of students at or above *Proficient* were in Massachusetts, Minnesota, Montana, and Ohio, which were not found to differ significantly from one another.

Figure 2.12: Cross-State Achievement Level Comparisons, Grade 4

Comparisons of percentage of students at or above *Proficient* in science for grade 4 public schools: 2000

Instructions: Read down the column directly under a jurisdiction name listed in the heading at the top of the figure. Match the shading intensity surrounding a jurisdiction's abbreviation to the key below to determine whether the percentage of students at or above *Proficient* in this jurisdiction is higher than, or lower than the jurisdiction in the column heading. For example, in the column under Michigan, the percentage of students in Michigan was lower than Massachusetts, all the states from Vermont through Oregon, and higher than the remaining states down the column.



Jurisdiction has statistically significantly higher percentage than the jurisdiction listed at the top of the figure.

No statistically significant difference detected from the jurisdiction listed at the top of the figure.

Jurisdiction has statistically significantly lower percentage than the jurisdiction listed at the top of the figure.

The between jurisdiction comparisons take into account sampling and measurement error and that each jurisdiction is being compared with every other jurisdiction. Significance is determined by an application of a multiple comparison procedure (see appendix A).

++Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see appendix A).

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools.
DoDDS: Department of Defense Dependents Schools (Overseas).

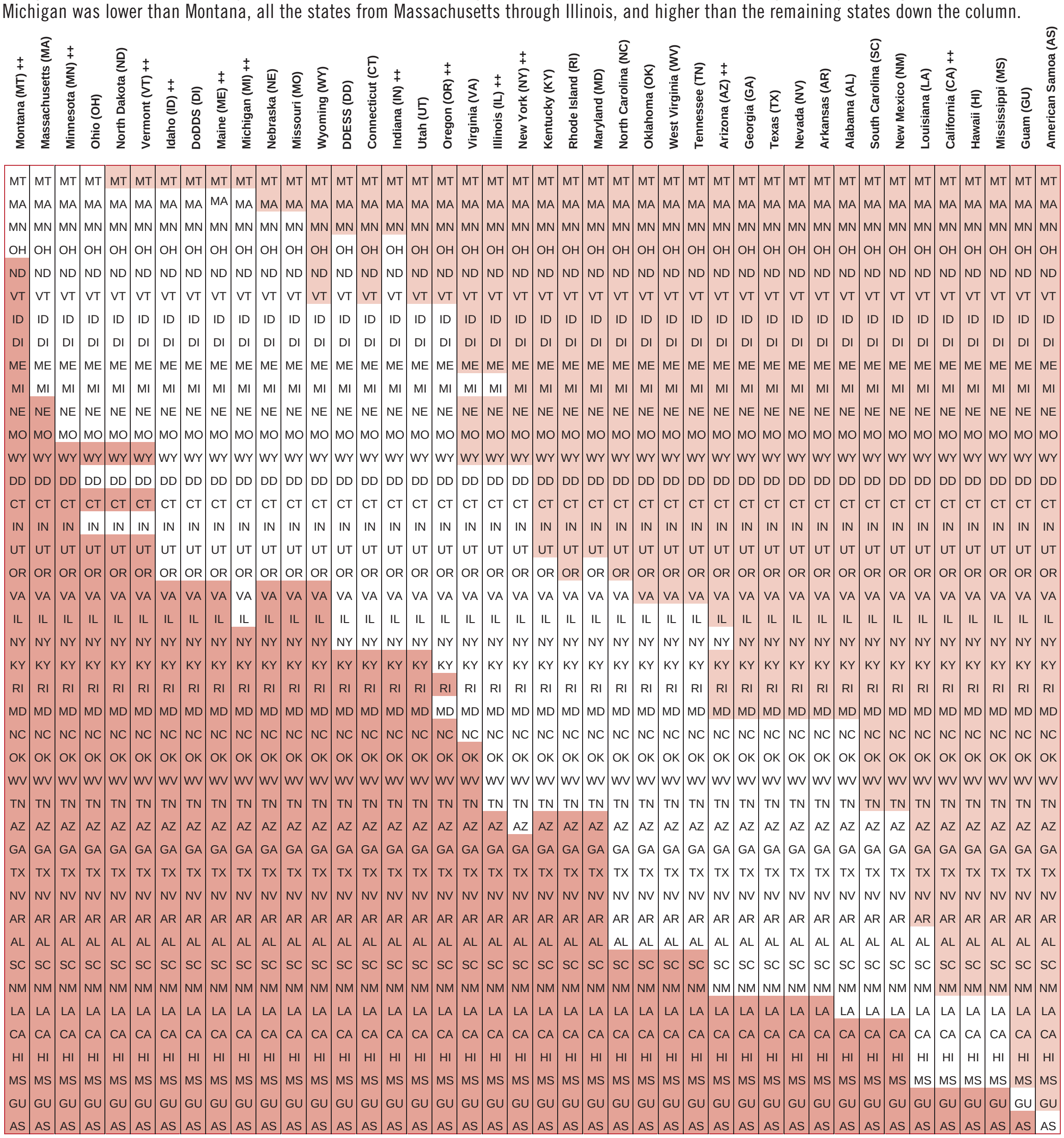
NOTE: Differences between states and jurisdictions may be partially explained by other factors not included in this figure.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.

Figure 2.13: Cross-State Achievement Level Comparisons, Grade 8

Comparisons of percentage of students at or above *Proficient* in science for grade 8 public schools: 2000

Instructions: Read down the column directly under a jurisdiction name listed in the heading at the top of the figure. Match the shading intensity surrounding a jurisdiction's abbreviation to the key below to determine whether the percentage of students at or above *Proficient* in this jurisdiction is higher than, or lower than the jurisdiction in the column heading. For example, in the column under Michigan, the percentage of students in Michigan was lower than Montana, all the states from Massachusetts through Illinois, and higher than the remaining states down the column.



Jurisdiction has statistically significantly higher percentage than the jurisdiction listed at the top of the figure.

No statistically significant difference detected from the jurisdiction listed at the top of the figure.

Jurisdiction has statistically significantly lower percentage than the jurisdiction listed at the top of the figure.

The between jurisdiction comparisons take into account sampling and measurement error and that each jurisdiction is being compared with every other jurisdiction. Significance is determined by an application of a multiple comparison procedure (see appendix A).

++Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see appendix A).

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools.
DoDDS: Department of Defense Dependents Schools (Overseas).

NOTE: Differences between states and jurisdictions may be partially explained by other factors not included in this figure.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 Science Assessment.