DOCUMENT RESUME

ED 414 202 SE 061 018

AUTHOR	Ballator, Nada; O'Sullivan, Christine Y.; Jerry, Laura
TITLE	NAEP 1996 Science State Report for Department of Defense Domestic Dependent Elementary and Secondary Schools, Grade 4. Findings from the National Assessment of Educational Progress.
INSTITUTION	Educational Testing Service, Washington, DC.; National Assessment of Educational Progress, Princeton, NJ.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-98-486
PUB DATE	1997-10-00
NOTE	127p.; For overall report, see ED 405 221. Produced "in collaboration with James E. Carlson, Audrey Champagne, Peggy Carr, John R. Donoghue, Will Pfeiffenberger, and Mistilina Sato." The name "Fiona Herr" appears as an author on the title page, but was later deleted by an errata page, which also re-sequenced the authors.
AVAILABLE FROM	National Library of Education, Office of Educational Research and Improvement, U.S. Dept. of Education, 555 New Jersey Ave., N.W., Washington, DC 20208-5574; phone: 800-424-1616 (limit one copy); also on NCES web site: http://nces.ed.gov/naep/96state
PUB TYPE	Information Analyses (070) -- Reports - Research (143)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Academic Achievement; Academic Standards; Educational Change; *Grade 4; Hands on Science; Intermediate Grades; *National Competency Tests; Problem Solving; *Science Education; Science Process Skills; Sex Differences; *Standardized Tests; *Student Evaluation; Tables (Data)
IDENTIFIERS	*Dependents Schools; National Assessment of Educational Progress; State Science Assessment (NAEP)

ABSTRACT

In 1990, the National Assessment of Educational Progress (NAEP) included a Trial State Assessment (TSA); for the first time in the NAEP's history, voluntary state-by-state assessments were made. The sample was designed to represent the 8th grade public school population in a state or territory. In 1996, 44 states, the District of Columbia, Guam, and the Department of Defense schools took part in the NAEP state science assessment program. The NAEP 1996 state science assessment was at grade 8 only, although grades 4,8, and 12 were assessed at the national level as usual. Both the domestic and overseas Department of Defense schools made special arrangements to assess their grade 4 students during the national science assessment. The results reported here are from the grade 4 assessment of the Domestic Dependents Elementary and Secondary Schools (DDESS). The 1996 state science assessment covered three major fields: earth, physical, and life sciences. In DDESS, 1,251 students in 39 public schools were assessed. This report describes the science proficiency of DDESS fourth-graders, compares their overall performance to students in the entire United States (using data from the NAEP national assessment), presents the average proficiency for the three major fields, and summarizes the performance of subpopulations (gender, race/ethnicity, parents' educational level, Title I participation, and free/reduced lunch program eligibility). To provide a context for the assessment data, participating students, their science teachers, and principals completed questionnaires which focused on: instructional content (curriculum coverage, amount of homework); delivery of science instruction (availability of resources, type); use of computers in science instruction; educational background of teachers; and conditions facilitating science learning (e.g., hours of television watched, absenteeism). On the NAEP fields of science scales that range from 0 to 300, DDESS students had an average proficiency of 154 compared to 148 throughout the United States. The average science scale score of males did not differ significantly from that of females in DDESS. However, the scores of both DDESS males and females were significantly higher than for males and females nationwide. At the fourth grade, White students in DDESS had an average science scale score that was higher than those of Black and Hispanic students. (SGE)

NATIONAL CENTER FOR EDUCATION STATISTICS

NAEP 1996 SCIENCE

Report for Department of Defense Domestic Dependent Elementary and Secondary Schools Grade 4

THE NATION'S REPORT CARD naep

U.S. DEPARTMENT OF EDUCATION OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT

Findings from the National Assessment of Educational Progress

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 a congressionally mandated project of the National Center for Education Statistics, the U.S. Department of Education. The Commissioner of Education Statistics 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

NAEP 1996 SCIENCE STATE REPORT

for

DEPARTMENT OF DEFENSE DOMESTIC DEPENDENT ELEMENTARY AND SECONDARY SCHOOLS Grade 4

Christine Y. O'Sullivan
Laura Jerry
Nada Ballator
Fiona Herr

In collaboration with
Audrey Champagne, Peggy Carr,
Will Pfeiffenberger, and Mistilina Sato

October 1997

U.S. Department of Education
Office of Educational Research and Improvement

Prepared by Educational Testing Service under a cooperative agreement with the National Center for Education Statistics.

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Ricky T. Takai
Acting Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr.
Commissioner

Education Assessment Group
Gary W. Phillips
Associate Commissioner

October 1997

SUGGESTED CITATION
O'Sullivan, C.Y., Jerry, L., Ballator, N., and Herr, F.
NAEP 1996 Science State Report for Department of Defense Domestic Dependent Elementary and Secondary Schools, Grade 4,
Washington, DC: National Center for Education Statistics, 1997.

FOR MORE INFORMATION
Contact:
Arnold A. Goldstein
202-219-1741

For ordering information on this report, write:

National Library of Education
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, D.C. 20208-5641

or call 1-800-424-1616 (in the Washington, DC, metropolitan area call 202-219-1651).

This report also is available on the World Wide Web: http://www.ed.gov/NCES/naep

The work upon which this publication is based was performed for the National Center for Education Statistics, Office of Educational Research and Improvement, by Educational Testing Service.

ERRATA

NAEP 1996 Science Report for Department of Defense Domestic Dependent Elementary and Secondary Schools Grade 4

Title page - The authors of this report are Nada Ballator, Christine Y. O'Sullivan and Laura Jerry. The report was produced in collaboration with James E. Carlson, Audrey Champagne, Peggy Carr, John R. Donoghue, Will Pfeiffenberger, and Mistilina Sato.

Citation page - The suggested citation for this report is as follows:

Ballator, N., O'Sullivan, C.Y., and Jerry, L.
NAEP 1996 Science State Report for Department of Defense Domestic Dependent Elementary and Secondary Schools, Grade 4, Washington, DC: National Center for Education Statistics, 1997.

ERRATA NOTICE

Date: December 29, 1997

To: Participants in the NAEP 1996 Science State Assessment

From: Nada Ballator
Center for the Assessment of Educational Progress at Educational Testing Service
1-800-223-0267

Re: Replacement pages attached for *NAEP 1996 Science State Reports*, correcting error in national and regional data in Table 6.2 and associated text

---

An error was recently discovered in the ***national and regional*** data presented in Table 6.2 of the 1996 science state reports. ***For all states and jurisdictions, the data are correct;*** however, incorrect national data made it necessary to recompute comparisons between state and national results. The error involved the student background item, "About how many books are in your home?" which is reported in the *NAEP 1996 Science State Report* in Table 6.2, as well as in the bullets comparing your jurisdiction with the nation.

Attached to this memo are the two corrected pages to insert into your printed reports. If you received camera-ready copy of the NAEP 1996 science state report, we have also enclosed pages for insertion there. The pages are for Chapter 6 in the section on "Literacy Materials in the Home" which includes Table 6.2; they contain revised comparisons to national data, and revised national and regional data in the table. We apologize for the publication of inaccurate data, and for the extra effort its correction will cause you.

DONE

The state science reports also appear on the NCES web site (**http://nces.ed.gov/naep**). All affected reports on the web were corrected on December 17. There is now a **Revised** logo beside the reports on the Index of Results and Summary Data web page (**http://nces.ed.gov/naep/rsdindex.shtml**) and on the Current Assessment Results web page (**http://nces.ed.gov/naep/naep1996.html**), and an **Errata Notice** containing a brief description of the repair on the NAEP 1996 Science State Reports web page (**http://nces.ed.gov/naep/96state/97499.shtml**).

Also on the web site, the student data tables for national science results for public schools have been revised. On the web page for NAEP 1996 Summary Data Tables, Student Data (**http://nces.ed.gov/naep/tables96/index.shtml**), you will see an **Errata Notice** describing the repair. Please alert anyone who may be using national 1996 science student data to this revision concerning the raw variable, "How many books are in your home," and the derived variable HOMEEN3, "Home environment - Articles (of 4) in home."

We very much regret the extra work that this error may have necessitated in your jurisdiction; we will redouble our efforts to prevent such things happening again.

Table of Contents

8

HIGHLIGHTS

Monitoring the performance of students in subjects such as science is a key concern of the citizens, policy makers, and educators who direct educational reform efforts. The 1996 National Assessment of Education Progress (NAEP) in science assesses the current level of science performance as a mechanism for informing education reform. This science assessment is the first to be constructed on a new framework.

The Department of Defense Education Activity (DoDEA) comprises schools in the domestic United States as well as schools attached to United States agencies overseas. The DoDEA domestic and overseas schools both participated in the 1996 science state assessment program at grade 8, and both jurisdictions also made special arrangements to assess their grade 4 students during the national science assessment. The results reported here are from the grade 4 assessment of the Domestic Dependents Elementary and Secondary Schools (DDESS). The results for fourth graders from the overseas Department of Defense Dependent Schools (DoDDS) are in a companion report.

What is NAEP?

The National Assessment of Educational Progress (NAEP), the "Nation's Report Card," is the only ongoing nationally representative assessment of what students in the United States know and can do in various academic subjects. Since 1969, NAEP assessments have been conducted with national samples of students in the subject areas of reading, mathematics, science, writing, and other fields. By making information on student performance available to policy makers, educators, and the general public, NAEP is an integral part of our nation's evaluation of the conditions and progress of education.

NAEP is a congressionally mandated project of the National Center for Education Statistics (NCES), U.S. Department of Education. Results are provided only for group performance. NAEP is forbidden by law to report results at an individual or school level.

In 1990 Congress authorized a voluntary state-by-state NAEP assessment. State-level assessments have taken place in mathematics (in 1990, 1992, and 1996), and reading in 1992 and 1994. In 1996, 44 states, the District of Columbia, Guam, and the DoDEA schools volunteered to take part in the NAEP State Assessment Program at grade 8. The results for each jurisdiction are reported in the *NAEP 1996 Science State Reports,* which are available in print and also on the NCES web site (http://www.ed.gov/NCES/naep).

NAEP 1996 Science Assessment

The framework for the science assessment was produced through a national consensus process by educators, administrators, assessment experts and curriculum specialists. The framework was designed to reflect current practices in science teaching. It called for the use of multiple-choice questions and constructed-response questions that required both short and extended responses. The constructed-response questions served as indicators of students' ability to know and integrate facts and scientific concepts, the ability to reason, and the ability to communicate scientific information. In the 1996 assessment, these constructed-response questions constituted nearly 80 percent of the total student response time. The NAEP 1996 assessment in science also included hands-on tasks that enabled students to demonstrate directly their knowledge and skills related to scientific investigation.

The 1996 science framework was structured according to a matrix that consisted of the three traditional fields of science (earth, physical, and life) crossed with three processes of knowing and doing science (conceptual understanding, scientific investigation, and practical reasoning). A central category encompassing the nature of science and the nature of technology was woven throughout the assessment, as was a themes category representing major ideas or key concepts that transcend scientific disciplines.

Students' science performance is summarized on the NAEP science scales, which range from 0 to 300 at each grade. While the scale score ranges are identical for grades 4, 8, and 12, the scales were derived independently at each grade. Scale scores on the grade 4 scale cannot imply anything about performance at grade 8 in the assessment.

Comparison of DDESS to the Nation

Table H.1 shows the distribution of science scale scores for the fourth-grade students attending DDESS schools in 1996. For this table and the others throughout this report, the results shown for Nation are from the national sample of public schools only.

- The average science scale score for fourth graders in DDESS was 154. This average was significantly higher than that for the nation (148).[1]

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE H.1

Distribution of Science Scale Scores for Grade 4 Students

	Average Scale Score	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
DDESS	154 (0.9)	116 (3.0)	134 (1.6)	155 (1.5)	175 (1.1)	190 (1.9)
Nation	148 (0.9)	103 (1.3)	127 (1.8)	151 (1.2)	172 (0.9)	188 (1.4)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

[1] Differences reported as significant are statistically different at the 95 percent confidence level. This means that with 95 percent confidence there is a real difference in average science scale score between the two populations of interest.

Major Findings for Student Subpopulations

The preceding section provided a view of the overall science performance of fourth-grade students in DDESS. It is also important to examine the average science scale scores of subgroups within the population. Typically, NAEP presents results for demographic subgroups such as those defined by gender, race/ethnicity, and parental education. In addition, in 1996 NAEP collected information on student participation in two federally funded programs — Title I and the free/reduced-price component of the National School Lunch Program.

The reader is cautioned against using NAEP results to make simple or causal inferences related to subgroup membership. Differences among groups of students are almost certainly associated with a broad range of socioeconomic and educational factors not discussed in NAEP reports and possibly not addressed by the NAEP assessment program.

Results related to gender and race/ethnicity are highlighted below. More complete results for the various demographic subgroups examined by the NAEP science assessment can be found in Chapters 2 and 4 of this report.

- The average science scale score of males (153) did not differ from that of females (154) in DDESS schools. However, the scores of both males and females were significantly higher than for males (149) and females (148) nationwide.

- At the fourth grade, White students in DDESS demonstrated an average science scale score (164) that was significantly higher than that of Black (143) and Hispanic students (144).

Finding a Context for Understanding Students' Science Performance

The science performance of students in DDESS may be better understood when viewed in the context of the environment in which students are learning. This educational environment is largely determined by school policies and practices, by characteristics of science instruction in the school, by home support for academics and other home influences, and by students' own views about science. Information about this environment is gathered by means of questionnaires completed by principals and teachers as well as questions answered by students as part of the assessment.

Because NAEP is administered to a sample of students that is representative of all fourth-grade students in the DDESS, NAEP results provide a view of the educational practices that may be useful for improving instruction and setting policy. However, despite the richness of context provided by the NAEP results, it is very important to note that NAEP data cannot establish a cause-and effect relationship between educational environment and students' scores on the NAEP science assessment.

School Science Education Policies and Practices[2]

- In DDESS, the percentage of fourth-grade students attending schools that reported science was a priority (42 percent) was not different from the percentage of fourth-grade students nationwide (42 percent).

- The percentage of fourth-grade public school students in DDESS who attended schools that reported having a district or state science curriculum that the school was expected to follow (87 percent) was not significantly different than the national percentage (92 percent).

- In DDESS, 68 percent of fourth graders attended schools that reported providing instruction in science every day. This percentage was greater than that of fourth graders across the nation (47 percent).

- A small percentage of students in DDESS had teachers who reported receiving all of the resources they needed for science instruction in DDESS (18 percent). This was higher than that of fourth-grade public school students nationwide (10 percent).

- In DDESS, 52 percent of the fourth-grade students were taught by teachers who reported that there was a curriculum specialist available to help or advise the teachers in science. This figure did not differ significantly from that of students across the nation (47 percent).

Science Classroom Practices[3]

- Less than half of the fourth-grade students in DDESS had science teachers who reported spending a lot of time on life science (40 percent), about one third reported spending a lot of time on earth science (32 percent), and about one third reported spending a lot of time on physical science (30 percent).

- In DDESS, 49 percent of the fourth graders had teachers who planned to emphasize heavily the students' knowledge of science facts and terminology. At the other extreme, 3 percent of the students had teachers who planned little to no emphasis on this topic.

- Teachers of 61 percent of the fourth-grade students reported that they placed heavy emphasis on developing science problem-solving skills. A small percentage of the students (3 percent) had teachers who reported spending little or no time addressing this topic.

- In terms of learning how to communicate ideas in science effectively, 34 percent of the fourth-grade students in DDESS had teachers who reported heavily emphasizing this ability for their students, while 7 percent of the students had teachers who reported giving little to no emphasis on this topic.

---

2 More detailed results related to school policies and practices can be found in Chapter 3 of this report, the *NAEP 1996 Science Report for Grade 4 DoDEA/DDESS*.

3 *Ibid.*: More detailed results related to classroom practices can be found in Chapter 4 of this report.

- In DDESS, 23 percent of fourth graders reported not spending any time on science homework in a typical week. By comparison, 21 percent spent one hour or more on their science homework each week.

Scientific Investigations[4]

- Of the fourth-grade students in DDESS, 84 percent had teachers who reported giving moderate to heavy emphasis on the development of data analysis skills. This percentage was significantly higher than that of students nationwide (65 percent).

- More than half of the fourth graders in DDESS had teachers who reported that their students performed hands-on activities or investigations in science once a week or more (63 percent).

Influences Beyond School That Facilitate Learning Science[5]

- The percentage of fourth graders in DDESS who reported watching six or more hours of television a day (20 percent) was not significantly different from the percentage for the nation (21 percent).

- In DDESS, 37 percent of fourth-graders agreed that science is useful for solving everyday problems. This was not significantly higher than for public school students in the nation (34 percent).

[4] *Ibid.*; More detailed results related to scientific investigations can be found in Chapter 5 of this report.
[5] *Ibid.*; More detailed results related to influences beyond school that facilitate learning science can be found in Chapter 6 of this report.

INTRODUCTION

Improving education is often seen as an important first step as the United States maps out a strategy to remain competitive in an increasingly technical global economy. At the 1996 Governors' Summit in Palisades, New Jersey, the President and the Governors reaffirmed the need to strengthen the nation's schools and to strive for world-class standards. Furthermore, in his 1997 State of the Union Address, President Clinton placed education center stage and called for states to commit to national standards that represent what all students must know to succeed in the knowledge-based economy of the twenty-first century.

In 1983, the National Commission on Excellence in Education issued a report entitled *A Nation at Risk: The Imperative for Education Reform* that was critical of education in the United States.[6] Interest in reform was also fueled by the publication of other reports and analyses[7] that pointed out the deficiencies of the educational system and how these could be rectified. Since then, organizations from the public and private sectors have assumed pivotal roles in providing support to state and local educational establishments as they seek to reform their educational systems[8] in areas such as the development of standards, revision of curricula, development of appropriate assessment techniques, and professional development. In addition to these activities, organizations such as the National Science Teachers Association and the American Association for the Advancement of Science have worked closely with the National Research Council to produce documents that help teachers interpret the National Science Education Standards[9] that were published in 1995. As the new millennium approaches, commitment to science reform continues.

Monitoring the performance of students in science is a key concern of the state and national policy makers and educators who direct educational reform efforts. To this end, the 1996 National Assessment of Educational Progress (NAEP) is an important source of information on what the nation's students know and can do in science.

---

[6] *A Nation at Risk: The Imperative for Education Reform.* (Washington, DC: National Commission on Excellence in Education, 1983).

[7] *Educating Americans for the 21st Century: A Report to the American People and the National Science Board.* (Washington, DC: National Science Board, Commission on Precollege Education in Mathematics, Science, and Technology, 1983).

[8] *Statewide Systemic Initiatives in Science, Mathematics, and Engineering.* (Arlington, VA: The National Science Foundation, 1995-1996); *Scope, Sequence, and Coordination of Secondary School Science. Volume I: The Content Core; Volume II: Relevant* (Washington, DC: National Science Teachers Association, 1992); *Benchmarks for Science Literacy.* (Washington, DC: Project 2061, American Association for the Advancement of Science, 1993); New Standards Project. (Washington, DC: National Research Council, 1995).

[9] *National Science Education Standards.* (Washington, DC: National Research Council, 1995).

What Was Assessed?

The science assessment was crafted to measure the content and skills specified in the science framework for the 1996 National Assessment of Educational Progress. Two organizing concepts underlie the science framework. First, scientific knowledge should be structured so as to make factual information meaningful. The way in which knowledge is structured should be influenced by the context in which the knowledge is being presented. Second, science performance depends on the knowledge of facts, the ability to integrate this knowledge into larger constructs, and the capacity to use the tools, procedures, and reasoning processes of science to develop an increased understanding to the natural world. Thus, the framework called for the NAEP 1996 science assessment to include the following:

- Multiple-choice questions that assess students' knowledge of important facts and concepts and that probe their analytical reasoning skills;

- Constructed-response questions that explore students' abilities to explain, integrate, apply, reason about, plan, design, evaluate, and communicate scientific information; and

- Hands-on tasks that probe students' abilities to use materials to make observations, perform investigations, evaluate experimental results, and apply problem-solving skills.

The core of the science framework is organized along two dimensions. The first dimension divides science into three major fields: earth, physical, and life sciences. The second dimension defines characteristic elements of knowing and doing science: conceptual understanding, scientific investigation, and practical reasoning. Each question in the assessment is categorized as measuring one of the elements of knowing and doing within one of the fields of science (e.g., scientific investigation in the context of earth science). The framework also contains two overarching domains — the nature of science and the organizing themes of science. The nature of science encompasses the historical development of science and technology, the habits of mind that characterize science, and the methods of inquiry and problem solving. It also includes the nature of technology — specifically, design issues involving the application of science to real-world problems and associated trade-offs or compromises. The themes of science include the notions of systems and their application in the scientific disciplines, models and their functioning in the development of scientific understanding, and patterns of change as they are exemplified in natural phenomena. A fuller description of the framework is provided in Appendix B.

Who Was Assessed?

<u>School and Student Characteristics</u>

Table I.1 provides demographic profiles of the fourth-grade students in DDESS and in the nation's public schools. These profiles are based on data collected from the DDESS students and schools participating in the 1996 national science assessment at grade 4. As described in Appendix A, the DDESS data and the national data are drawn from separate samples.

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE I.1

Profile of Grade 4 Students in DDESS and the Nation

Demographic Subgroups		Public Schools Percentage
RACE/ETHNICITY		
DDESS	White	46 (1.2)
	Black	27 (1.1)
	Hispanic	20 (0.9)
	Asian/Pacific Islander	3 (0.4)
	American Indian	3 (0.6)
Nation	White	67 (0.7)
	Black	15 (0.4)
	Hispanic	13 (0.6)
	Asian/Pacific Islander	3 (0.2)
	American Indian	2 (0.2)
PARENTS' EDUCATION		
DDESS	Did not finish high school	1 (0.3)
	Graduated from high school	12 (0.9)
	Some education after high school	11 (0.8)
	Graduated from college	39 (1.5)
	I don't know	37 (1.1)
Nation	Did not finish high school	5 (0.4)
	Graduated from high school	14 (0.8)
	Some education after high school	8 (0.5)
	Graduated from college	39 (1.6)
	I don't know	35 (1.0)
GENDER		
DDESS	Male	52 (1.5)
	Female	48 (1.5)
Nation	Male	50 (0.7)
	Female	50 (0.7)
TITLE 1		
DDESS	Participated	0 (****)
	Did not participate	100 (****)
Nation	Participated	24 (2.0)
	Did not participate	76 (2.0)
FREE/REDUCED-PRICE LUNCH		
DDESS	Eligible	35 (1.3)
	Not eligible	37 (1.3)
	Information not available	29 (0.8)
Nation	Eligible	39 (2.2)
	Not eligible	54 (2.4)
	Information not available	8 (2.0)

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

Schools and Students Assessed

Table I.2 summarizes participation[10] data for schools and students sampled in DDESS for the 1996 Assessment in science at grade 4.

In DDESS, 39 schools participated in the 1996 fourth-grade science assessment. These numbers include participating substitute schools that were selected to replace some of the nonparticipating schools from the original sample. The weighted school participation rate after substitution in 1996 was 100 percent, which means that the fourth-grade students in this sample were directly representative of 100 percent of all the fourth-grade students in DDESS.

In each school, a random sample of students was selected to participate in the assessment. In DDESS in 1996, on the basis of sample estimates, 0 percent of the fourth-graders were classified as students with limited English proficiency (LEP). In addition, 10 percent of fourth graders had an Individual Education Plan (IEP). An IEP is a plan written for a student who has been determined to be eligible for special education. The IEP typically sets forth goals and objectives for the student and describes a program of activities and/or related services necessary to achieve the goals and objectives. A student with an IEP may be classified as SD (student with disabilities).

[10] For a detailed discussion of the NCES guidelines for sample participation, see Appendix A of this report or the *Technical Report of the NAEP 1996 State Assessment Program in Science*. (Washington, DC: National Center for Education Statistics, 1997).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE I.2

School and Student Participation in DDESS at Grade 4

	Public Schools
SCHOOL PARTICIPATION	
Weighted school participation rate before substitution	100%
Weighted school participation rate after substitution	100%
Number of schools originally sampled	39
Number of schools not eligible	0
Number of schools in original sample participating	39
Number of substitute schools provided	0
Number of substitute schools participating	0
Total number of participating schools	39
STUDENT PARTICIPATION	
Weighted student participation rate after makeups	96%
Number of students selected to participate in the assessment	1404
Number of students withdrawn from the assessment	139
Percentage of students who were of Limited English Proficiency	0%
Percentage of students excluded from the assessment due to Limited English Proficiency	0%
Percentage of students who had an Individualized Education Plan	10%
Percentage of students excluded from the assessment due to Individualized Education Plan status	6%
Number of students to be assessed	1310
Number of students assessed	1251
Overall weighted response rate	96%

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

Schools were permitted to exclude certain students from the assessment, provided that the following criteria were met. To be excluded, a student had to be categorized as LEP or had to have an IEP *and* (in either case) be judged incapable of participating in the assessment. The intent was to assess all selected students; therefore, all selected students who were capable of participating in the assessment should have been assessed. However, schools were allowed to exclude those students who, in the judgment of school staff, could not meaningfully participate. The NAEP guidelines for exclusion are intended to assure uniformity of exclusion criteria from school to school. Note that some students classified as LEP and some students having an IEP were deemed eligible to participate and were included in the assessment. In DDESS, no students were excluded from the assessment because they were categorized as LEP, and those excluded due to the specifications of their IEP represented 6 percent of the population in grade 4.

In DDESS, 1,251 fourth-grade students were assessed in 1996. The weighted student participation rate was 96 percent. This means that the sample of fourth-grade students who took part in the assessment was directly representative of 96 percent of the eligible DDESS student population (that is, all students from the population represented by the participating schools, minus those students excluded from the assessment). The overall weighted response rate (school rate times student rate) was 96 percent This means that the sample of students who participated in the assessment was directly representative of 96 percent of the eligible fourth-grade DDESS population.

In accordance with standard practice in survey research, the results presented in this report were based on calculations that incorporate adjustments for the nonparticipating schools and students. Hence, the final results derived from the sample provide estimates of the science performance for the full population of eligible fourth-grade students in DDESS schools. However, in instances where nonparticipation rates are large, these nonparticipation adjustments may not adequately compensate for the missing sample schools and students.

In order to guard against potential nonparticipation bias in published results, the National Center for Education Statistics (NCES) has established minimum participation levels as a condition for the publication of 1996 results. NCES also established additional guidelines addressing four ways in which nonparticipation bias could be introduced into a jurisdiction's published results (see Appendix A). In 1996, DDESS met minimum participation levels at grade 4 and met all other established NCES participation guidelines.

In the analysis of student data and reporting of results, nonresponse weighting adjustments have been made at both the school and student level, with the aim of making the sample of participating students as representative as possible of the entire eligible fourth-grade population. For details of the nonresponse weighting adjustment procedures, see the *Technical Report of the NAEP 1996 State Assessment Program in Science*.[11]

[11] In 1996, the State program assessed science at grade 8. DoDEA schools (DDESS and DoDDS) participated in the state program at grade 8, but also made special arrangements to assess their grade 4 students, as reported here.

Reporting NAEP Science Results

The NAEP Science Scale

The NAEP 1996 science assessment spans the broad field of science in each of the grades assessed. Because of the survey nature of the assessment and the breadth of the domain, each student participating cannot be expected to answer all the questions in the assessment since this would impose an unreasonable burden on students and their schools. Thus, each student was administered a portion of the assessment, and data were combined across students to report on the achievement of fourth graders and on the achievement of subgroups of students (e.g., subgroups defined by gender or parental education).

Student responses to the assessment questions were analyzed to determine the percentage of students responding correctly to each multiple-choice question and the percentage of students achieving each of the score categories for constructed-response questions. Item response theory (IRT) methods were used to produce scales that summarized results for each of the three fields of science (e.g., earth, physical, and life) at each grade level. An overall composite scale also was developed at each grade by weighting the separate field of science scales based on its relative importance in the NAEP science framework. Results presented in this report are based on this overall composite scale, which ranges from 0 to 300.

The use of separate grade-specific reporting scales for the science assessment is consistent with the National Assessment Governing Board's 1993 policy that future NAEP assessments be developed using within-grade frameworks and that scaling be carried out within grade. The ranges of the science scales (from 0 to 300) differ by design from the 0-to-500 reporting scales used in other NAEP subject areas and were chosen to minimize confusion with other common test scales and to discourage inappropriate cross-grade comparisons. (Additional details of the scaling procedures can be found in Appendix C of this report and in the forthcoming *NAEP 1996 Technical Report*).

Science Achievement Levels

A companion report, being issued by the National Assessment Governing Board, will present the NAEP 1996 science results in terms of achievement levels. As authorized by the NAEP legislation and adopted by the National Assessment Governing Board, the achievement levels are based on the Board's judgments about what are reasonable performance expectations for students on the NAEP 1996 science assessment. The achievement levels for the NAEP 1996 science assessment were adopted on an interim basis, indicating that they may be revised when other information becomes available, such as the fourth and twelfth grade results from the Third International Mathematics and Science Study (TIMSS).

Interpreting NAEP Results

This report describes science performance for fourth graders and compares the results for various groups of students within that population — for example, those who have certain demographic characteristics or who responded to a specific background question in a particular way. The report examines the results for individual demographic groups and for individual background questions. It does not include an analysis of the relationships among combinations of these subpopulations or background questions.

Because the percentages of students in these subpopulations and their average science scale scores are based on samples, rather than on the entire population of fourth graders in a jurisdiction, the numbers reported are necessarily *estimates*. As such, they are subject to a measure of uncertainty, reflected in the *standard error* of the estimate. When the percentages or average scale scores of certain groups are compared, it is essential to take the standard error into account, rather than to rely solely on observed similarities or differences. Therefore, the comparisons discussed in this report are based on *statistical tests* that consider both the magnitude of the difference between the means or percentages and the standard errors of those statistics.

The statistical tests determine whether the evidence, based on the data from the groups in the *sample*, is strong enough to conclude that the averages or percentages are really different for those groups in the *population*. If the evidence is strong (i.e., the difference is statistically significant), the report describes the group averages or percentages as being different (e.g., one group performed *high than* or *lower than* another group) — regardless of whether the sample averages or sample percentages appear to be about the same or not. If the evidence is not sufficiently strong (i.e., the difference is not significant), the averages or percentages are described as being not significantly different — again, regardless of whether the sample averages or sample percentages appear to be about the same or widely discrepant. Rather than relying on the apparent magnitude of the difference between sample averages or percentages, the reader is cautioned to rely on the results of the statistical tests to determine whether those sample differences are likely to represent actual differences between the groups in the population. The statistical tests and the Bonferroni procedure, which is used when more than two groups are being compared, are discussed in greater detail in Appendix A.

In addition, some of the percentages reported in the text of the report are given qualitative descriptions (e.g., relatively few, about half, almost all, etc.). The descriptive phrases used and the rules used to select them are also described in Appendix A.

How Is This Report Organized?

The *NAEP 1996 Science Report for DDESS at Grade 4* is based on a computer-generated report that describes the science performance of fourth-grade students in the DDESS and the nation. A separate report describes additional fourth-grade science assessment results for DoDDS and for the nation[12] This report consists of four sections:

- An **Introduction** provides background information about what was assessed, who was sampled, and how the results are reported.

- **Part One** shows the distribution of science scale score results for fourth-grade students in DDESS and the nation.

- **Part Two** relates fourth-grade public school students' science scale scores to contextual information about school characteristics, instruction, and home support for science in DDESS schools and the nation. In addition, Chapter 5 discusses student results of the hands-on tasks.

- Several **Appendices** support the results discussed in the report:

Appendix A	Reporting NAEP 1996 Science Results
Appendix B	The NAEP 1996 Science Assessment
Appendix C	Technical Appendix
Appendix D	Teacher Preparation

Other Reports of NAEP 1996 Science Results

The following related reports may be of interest to the reader:

- *Cross-State Data Compendium for the 1996 Grade 8 Science Assessment*
- *Technical Report of the NAEP 1996 State Assessment Program in Science*
- *NAEP 1996 Science Report Card for the Nation and the States*
- *The NAEP 1996 Technical Report*

There are plans for several additional reports to appear in late 1997 and early 1998. These reports will contain sample questions with examples of student work, NAEP results related to policies and practices in schools and classrooms in the United States, and information from the special components of the 1996 NAEP, including the advanced science assessment and the hands-on exercises.

[12] O'Sullivan, C.Y., C.M. Reese, and J. Mazzeo. *NAEP 1996 Science Report Card for the Nation and the States.* (Washington, DC: National Center for Education Statistics, 1997).

PART ONE

Science Scale Score Results

The following chapters describe the average science scale scores of fourth-grade students in DDESS. As described in the Introduction, the NAEP science scale is a composite of the three major fields of science: earth, physical, and life. Student performance is generally reported on this composite scale, thus reflecting average student scores across the three fields. The composite science scale ranges from 0 to 300. Student performance may be summarized on separate NAEP fields of science scales that also range from 0 to 300.

This part of the report has two chapters. Chapter 1 compares the overall science performance of students in DDESS to the nation and has a table showing students' average scale score distributions for the three fields of science. Chapter 2 summarizes science performance for subpopulations of public school students as defined by gender, race/ethnicity, parental education, participation in Title I services and programs, and eligibility for the free/reduced-price lunch component of the National School Lunch Program (NSLP).

The NAEP 1996 assessment in science is the first developed using a new framework, described in Appendix B. The scale developed to report results from the 1996 science assessment is a within-grade scale comprised of three fields of science scales. Appendix A describes reporting on the scale, and Appendix C describes the construction of the scale.

Item Maps

Students' performance is summarized on the NAEP science scale, ranging from 0 to 300. Sample questions are shown in Figure 1.1 illustrating the range of performance on the NAEP science scale for grade 4. Each question is one that is likely to be answered correctly by a student whose score is at or near the given percentile.

To illustrate the range of performance in more detail, questions from the assessment were "mapped" onto a 0 to 300 scale, as in Figure 1.2. The item map is a visual representation of the scale showing selected questions in positions corresponding to their difficulty. The item map shows which questions a student of any particular ability is likely to answer correctly. The position of the question on the scale represents a dividing line. Students who attained scores greater than the score corresponding to the question's difficulty are likely to answer it correctly, while students with scores below that degree of difficulty are less likely to answer it correctly.

More specifically, students who scored below the scale score associated with a particular question had less than a 65 percent probability of earning a given amount of credit on a constructed-response question or less than a 74 percent probability of correctly answering a multiple-choice question. A small proportion of these students — those near but below the question's position on the scale — may be more likely than not to answer the question correctly (between 50 and 65 or 74 percent). Such students are not considered "able" to answer the question, since they have not achieved sufficient consistency in their responses.

This discussion and the item map illustrations refer to fourth-grade students in the national assessment, whose scores may not resemble those of fourth-grade students in DDESS.

THE NATION'S REPORT CARD naep 1996 State Assessment

FIGURE 1.1

Sample Questions Likely to Be Answered Correctly by Grade 4 Students At or Near Selected Percentiles

Percentile	Question
10th	*Identify items that conduct electricity.* (105)
25th	Read the level of a liquid in a graduated cylinder. (129)
50th	*Infer the function of teeth from diagrams showing their structure.* (152)
75th	*Explain the impact of fish death on an ecosystem.* (173)
90th	*Explain why Earth never runs out of water.* (192)

The value in parentheses represents the scale score attained by students who had a 65 percent probability of reaching a given level on a constructed-response question (above, in italic type) or a 74 percent probability of correctly answering a 4-option multiple choice question (above, in regular type).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Figure 1.2 is an item map for grade 4.[13] Multiple-choice questions are shown in regular type; constructed-response questions are in italic type.[14] An example of how to interpret the item map may be helpful. In this figure, a multiple-choice question involving interpreting a graph maps at the 137 point on the scale. This means that fourth-grade students with science scale scores at or above 137 are likely to answer this question correctly — that is, they have at least a 74 percent chance of doing so.[15] Put slightly differently, this question is answered correctly by at least 74 of every 100 students scoring at or above the 137 scale-score level. Note that this does not mean that students at or above the 137 scale score always answer the question correctly or that students below the 137 scale score always answer it incorrectly.

As another example, consider the constructed-response question that maps at a scale score of 192. This question concerns the supply of water on Earth. Scoring of responses to this question allows for partial credit by using a three-level scoring guide. Mapping a question at the 192 scale score indicates that at least 65 percent of the students performing at or above this point achieved a score of 3 ("Complete") on the question. Among students with lower scores, fewer than 65 percent gave complete responses to the question.

[13] Details on the procedures used to develop the item map are provided in the forthcoming *NAEP 1996 Technical Report*. The procedures are similar to those used in past NAEP assessments.

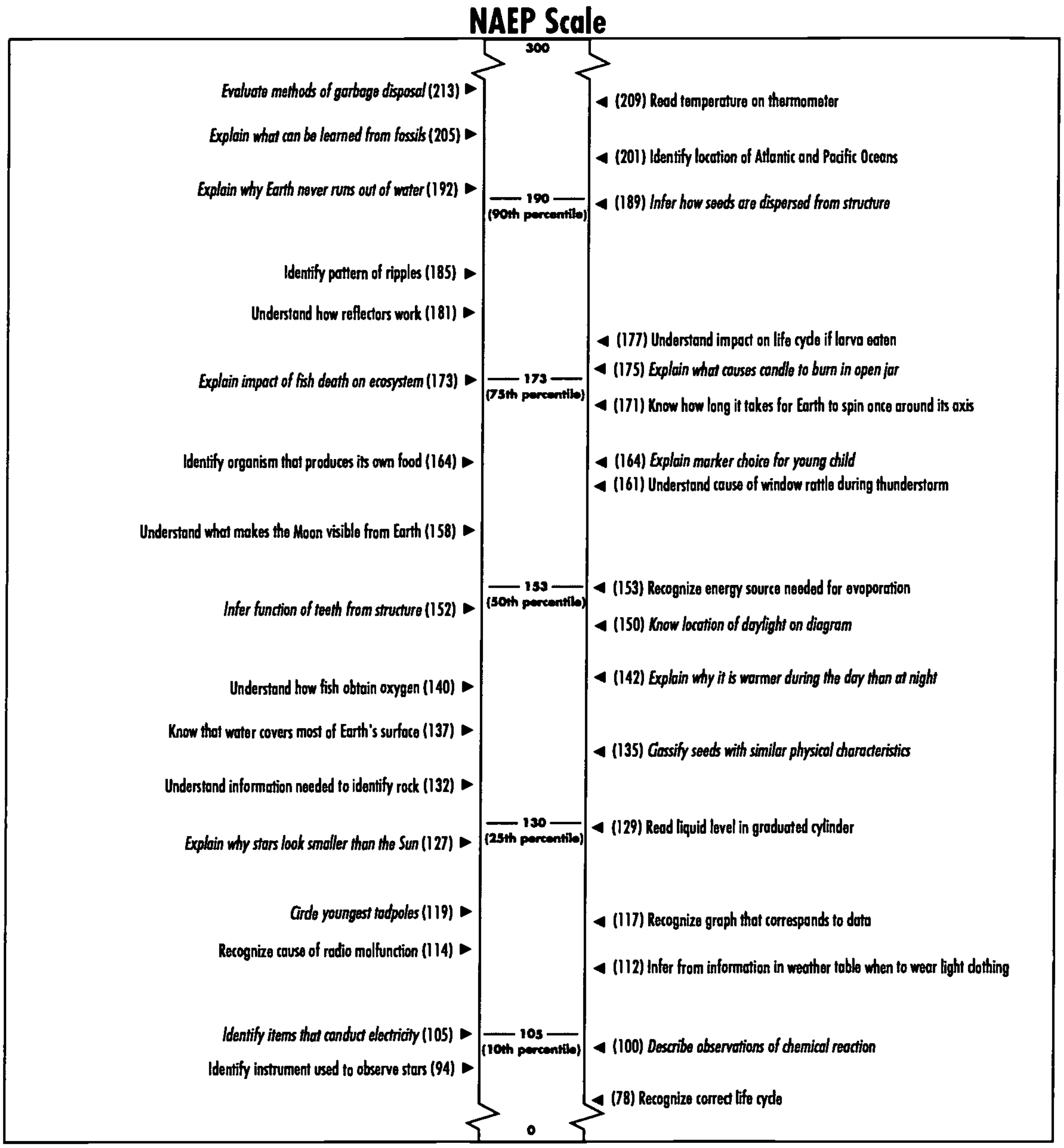
[14] The placement of constructed-response questions is based on (1) the "mapping" of a score of 3 on a 3-point scoring guide for short constructed-response questions and (2) the "mapping" of a score of at least 3 on a 4-point scoring guide and a score of at least 4 on a 5-point scoring guide for extended constructed-response questions.

[15] For constructed-response questions, a criterion of 65 percent was used. For multiple-choice questions, the criterion was 74 percent. The use of a higher criterion for multiple-choice questions reflected the students' ability to "guess" the correct answer from among the alternatives.

THE NATION'S REPORT CARD naep 1996 State Assessment

FIGURE 1.2 — GRADE 4

Map of Selected Questions on the NAEP Science Scale for Grade 4



NOTE: Position of questions is approximate and an appropriate scale range is displayed for grade 4. Italic type indicates a constructed-response question. Regular type denotes a multiple-choice question.

Each grade 4 science question was mapped onto the NAEP 0-to-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 reaching a given score level on a constructed-response question or a 74 percent probability of correctly answering a 4-option multiple-choice question. Only selected questions are presented. Percentiles of scale score distribution are referenced on the map.

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

CHAPTER 1

Science Scale Score Results for Fourth-Grade Students

To remain competitive in the global economy, a technologically and scientifically literate citizenry is required. As a result, reform in science and mathematics education in the United States has gained increasing attention. The 1983 publication *A Nation At Risk: The Imperative for Educational Reform* called for overall reform of the United States education system, with heavy emphasis placed on mathematics and science.[16] The National Goals Panel was convened in 1989 to further focus attention on education reform. In 1991 the National Science Foundation's Statewide Systemic Initiative began awarding grants to support state reform in K-12 mathematics and science instruction.[17] During the 1990s many states have been developing standards for science curriculum, teaching, and assessment using guidance from reform efforts such as the American Association for the Advancement of Science's *Project 2061*, the National Science Teachers Association's *Scope, Sequence and Coordination of High School Science*, and the recently published National Research Council's *National Science Education Standards*.[18] A reaffirmation of the United States' goal for world-class standards in education was made at the 1996 Governors' Summit in Palisades, NJ. These efforts all address ways to produce innovative science curricula aimed at improving national scientific literacy. As a means of informing the progress of such reform, the U.S. Department of Education supports programs geared toward assessing the current level of science knowledge and skills including the Third International Mathematics and Science Study (TIMSS),[19] administered in 1995, and the 1996 National Assessment of Educational Progress (NAEP) in science.

The NAEP 1996 state science assessment at grade 8 was the first time science had been assessed at the state level. It continued the state-level component begun in 1990 with the NAEP Trial State Assessment (TSA). The NAEP 1996 assessment in science had 47 participating

[16] *A Nation at Risk: The Imperative for Educational Reform.* (Washington, DC: National Commission on Excellence in Education, 1983).

[17] *Statewide Systemic Initiative.* (Washington, DC: National Science Foundation, 1990).

[18] *Science for All Americans: A Project 2061 Report on Literacy Goals in Science, Mathematics and Technology.* (Washington, DC: American Association for the Advancement of Science, 1989); *Scope, Sequence and Coordination of High School Science.* (Washington, DC: National Science Teachers Association, 1995); *National Science Education Standards.* (Washington, DC: National Research Council, 1996).

[19] The Third International Mathematics and Science Study was conducted in 1994 in the southern hemisphere and in 1995 in the northern hemisphere.

jurisdictions,[20] making it the largest NAEP state assessment so far. Results were reported for 46 of the 47 participating jurisdictions. The DoDEA schools participated in the science assessment at grade 8, and also made special arrangements for participation in the assessment at grade 4, although only the national program assessed students at that level.

The science framework for the 1996 National Assessment of Educational Progress[21] was developed through a consensus process involving educators, policy makers, business people, assessment experts, and curriculum specialists. The 1996 NAEP science assessment included multiple-choice questions, constructed-response exercises, and (for the first time) hands-on tasks. Because the 1996 assessment was based on an essentially new framework, it is not possible to compare results from the 1996 assessment with those from the previous NAEP science assessment in 1990.

Table 1.1 shows the distribution of science scale scores for fourth-grade students attending DDESS schools and public schools in the nation.

- The average science scale score in DDESS was 154. This average was higher than that for the nation (148).[22]

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 1.1

Distribution of Science Scale Scores for Grade 4 Students.

	Average Scale Score	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
DDESS	154 (0.9)	116 (3.0)	134 (1.6)	155 (1.5)	175 (1.1)	190 (1.9)
Nation	148 (0.9)	103 (1.3)	127 (1.8)	151 (1.2)	172 (0.9)	188 (1.4)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

[20] *Jurisdiction* refers to states, territories, the District of Columbia, and the Department of Defense Education Activities (DoDEA) domestic and overseas schools.

[21] *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1995).

[22] Differences reported as significant are statistically different at the 95 percent confidence level. This means that with 95 percent confidence there is a real difference in the average science scale score between the two populations of interest.

Performance in the NAEP Fields of Science Content Areas

The core of the science framework is organized along two dimensions. The first divides science into three major fields: earth, physical, and life. The second dimension defines characteristic elements of knowing and doing science: conceptual understanding, scientific investigation, and practical reasoning. Each question is categorized as measuring one of the elements of knowing and doing within one of the fields of science.

Table 1.2 shows the distribution of scale scores for each of the three fields of science for DDESS and the nation. Appendix B describes the three fields of science in more detail, and Appendix C contains a discussion of the scaling procedures used to develop the three fields of science scales and the composite NAEP science scale.

- The performance of students in DDESS was higher than that of students nationwide in the fields of earth science, physical science, and life science described in the framework for the assessment.

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 1.2

Distribution of Science Scale Scores for Grade 4 Students by Fields of Science

	Average Scale Score	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
Earth Science						
DDESS	153 (1.0)	111 (2.4)	133 (2.0)	155 (1.2)	176 (1.9)	193 (2.1)
Nation	148 (1.0)	101 (1.7)	127 (1.4)	151 (1.0)	173 (1.0)	191 (1.8)
Physical Science						
DDESS	154 (1.3)	108 (2.1)	132 (2.7)	156 (1.6)	178 (1.4)	196 (3.2)
Nation	148 (1.1)	101 (2.0)	125 (1.9)	150 (1.5)	172 (1.2)	191 (1.5)
Life Science						
DDESS	154 (1.0)	113 (2.2)	133 (2.1)	156 (1.5)	176 (1.9)	193 (2.1)
Nation	148 (1.0)	101 (2.2)	126 (1.5)	151 (1.1)	173 (1.0)	192 (1.7)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

BEST COPY AVAILABLE

CHAPTER 2

Science Scale Score Results for Fourth-Grade Students by Subpopulations

The previous chapter provided a view of the overall science performance of fourth-grade students in DDESS and the nation. It is also important to examine the average performance of subgroups since past NAEP assessments in science, as well as in other academic subjects, have shown substantial differences among groups defined by gender, racial/ethnic background, parental education, and other demographic characteristics.[23] A key contribution of NAEP to the ongoing conversations concerning education reform is the ability to monitor the performance of subgroups of students in academic achievement.

The NAEP 1996 state assessment in science provides performance information for subgroups of fourth graders in DDESS and the nation. In addition to the more typical demographic subgroups defined by gender, race/ethnicity and parental education, the 1996 assessment also collected information on two federally funded programs — student participation in Title I programs and services, and student eligibility for the free or reduced-price component of the National School Lunch Program (NSLP).

A description of the subgroups and how they are defined is presented in Appendix A. The reader is cautioned against making simple or causal inferences related to the performance of various subgroups of students or about the effectiveness of the NSLP or Title I programs, because average performance differences between two groups of students may be due in part to socioeconomic or other factors. For example, differences observed among racial/ethnic subgroups are almost certainly associated with a broad range of socioeconomic and educational factors not discussed in this report and possibly not addressed by the NAEP assessment program.[24] Similarly, differences in performance between students participating in Title I programs and students who are not does not account for the initial performance level of the students prior to placement in Title I programs or differences in course content and emphasis between the two groups.

[23] Jones, L.R., I.V.S. Mullis, S.A. Raizen, I.R. Weiss, and E.A. Weston. *The 1990 Science Report Card: NAEP's Assessment of Fourth, Eighth, and Twelfth Graders*. (Washington, DC: National Center for Education Statistics, 1992); Campbell, J.R., C.M. Reese, C. O'Sullivan, and J.A. Dossey. *NAEP 1994 Trends in Academic Progress*. (Washington, DC: National Center for Education Statistics, 1996).

[24] Investigating data from other sources may prove helpful; for example:
U.S. Department of Education . *Schools and Staffing in the United States: A Statistical Profile, 1993-94*. (Washington, DC: National Center for Education Statistics, 1996). URL: http://www.ed.gov/NCES/surveys/sass.html.

Gender

Previous NAEP results for science have shown a significant difference in the average scale scores of male and female eighth graders, with males consistently having higher scale scores.[25] As shown in Table 2.1, the NAEP 1996 state science assessment results for fourth graders in DDESS are not consistent with those general findings for the older students.

- The average science scale score of males (153) did not differ significantly from that of females (154) in DDESS. However, both male and female students in DDESS had higher average scores than their counterparts for the nation (149 for males and 148 for females).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 2.1

Distribution of Science Scale Scores for Grade 4 Public School Students by Gender

	Average Scale Score	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
Male						
DDESS	153 (1.6)	115 (3.0)	135 (1.5)	155 (1.5)	174 (1.2)	188 (2.7)
Nation	149 (1.0)	103 (1.2)	127 (1.9)	152 (1.4)	173 (0.9)	190 (1.4)
Female						
DDESS	154 (1.3)	117 (2.3)	134 (3.3)	155 (3.8)	176 (4.8)	192 (2.1)
Nation	148 (1.0)	103 (1.7)	128 (1.8)	150 (1.4)	170 (1.5)	187 (1.7)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Race/Ethnicity

As part of the background questions administered with the NAEP 1996 science assessment, students were asked to identify the racial/ethnic subgroup that best describes them. The five mutually exclusive categories were White, Black, Hispanic, Asian or Pacific Islander, and American Indian or Alaskan Native.

Findings from previous NAEP science assessments have shown that racial/ethnic differences exist in science performance.[26] However, when interpreting differences in subgroup performance, confounding factors related to socioeconomic status, home environment, and

[25] Campbell, J.R., K. E. Voelkl, and P. L. Donahue. *NAEP 1996 Trends in Academic Progress.* (Washington, DC: National Center for Education Statistics, 1997); Jones, L.R., I.V.S. Mullis, S.A. Raizen, I.R. Weiss, and E.A. Weston. *The 1990 Science Report Card: NAEP's Assessment of Fourth, Eighth, and Twelfth Graders.* (Washington, DC: National Center for Education Statistics, 1992).

[26] *Ibid.*

educational opportunities available to students need to be considered.[27] The distribution of fourth-grade science scale scores for DDESS and the nation by race/ethnicity are shown in Table 2.2.[28]

- White students in DDESS demonstrated an average science scale score (164)that was higher than that of Black (143) or Hispanic (144) DDESS students.

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 2.2

Distribution of Science Scale Scores for Grade 4 Public School Students by Race/Ethnicity

	Average Scale Score	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
White						
DDESS	164 (1.1)	129 (3.6)	148 (1.9)	166 (1.6)	182 (1.5)	197 (1.9)
Nation	158 (1.0)	121 (1.7)	140 (1.6)	159 (0.8)	178 (1.0)	193 (1.2)
Black						
DDESS	143 (1.6)	110 (3.9)	126 (3.7)	144 (2.0)	163 (0.8)	178 (3.3)
Nation	123 (1.9)	81 (2.6)	101 (1.7)	124 (2.5)	145 (2.1)	163 (2.1)
Hispanic						
DDESS	144 (2.4)	103 (4.8)	125 (5.2)	145 (4.2)	164 (1.2)	181 (2.2)
Nation	126 (1.7)	82 (3.6)	104 (2.5)	129 (1.6)	150 (1.6)	167 (2.5)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). The Asian/Pacific Islander Students (3%) and the American Indian Students (3%) were of insufficient sample size to permit reliable estimates of scores or standard errors.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

BEST COPY AVAILABLE

[27] McKenzie, F.D. "Educational Strategies for the 1990s," in *The State of Black America 1991*. (New York, NY: National Urban League, 1991).
[28] Results are reported for racial/ethnic subgroups meeting established sample size requirements (see Appendix A).

Students' Reports of Parents' Highest Education Level

Students were asked to indicate the level of education completed by each parent. Four levels of education were identified: did not finish high school, graduated from high school, some education after high school, and graduated from college. A choice of "I don't know" was also available. For this analysis the highest education level reported for either parent was used.

In general, results show that with each increment in reported parental education, student performance increases significantly. In reviewing these results, it is important to note that, nationally, approximately 10 percent of fourth graders did not know the level of education that either of their parents had completed. For public school students in DDESS, this percentage was 8 percent. Despite the fact that some research has questioned the accuracy of student-reported data from similar groups of students,[29] past NAEP assessments in science, as well as other subject areas, have found that student-reported level of parental education exhibits a consistently positive relationship with student performance on the assessments.[30] Other research has corroborated NAEP findings.[31]

Table 2.3 shows the results for fourth-grade public school students reporting that neither parent graduated from high school, at least one parent graduated from high school, at least one parent had received some education after high school, at least one parent graduated from college, or that they did not know their parents' highest education level. The following pertains to those students who reported knowing the educational level of one or both parents.

- The average science scale score of students in DDESS who reported that at least one parent graduated from high school (145) was significantly lower than that of students who reported that at least one parent had some education after high school (156), or that at least one parent graduated from college (160).

[29] Looker, E.D. "Accuracy of Proxy Reports of Parental Status Characteristics," in *Sociology of Education*, 62(4), pp. 257-276, 1989.
[30] Jones, L.R., I.V.S. Mullis, S.A. Raizen, I.R. Weiss, and E.A. Weston. *The 1990 Science Report Card: NAEP's Assessment of Fourth, Eighth, and Twelfth Graders.* (Washington, DC: National Center for Education Statistics, 1992); Campbell, J.R., P.L. Donahue, C.M. Reese, and G.W. Phillips. *NAEP 1994 Reading Report Card for the Nation and the States.* (Washington, DC: National Center for Education Statistics, 1996); Reese, C. M., K. E. Miller, J. Mazzeo, and J. A. Dossey. *NAEP 1996 Mathematics Report Card for the Nation and the States.* (Washington, DC: National Center for Education Statistics, 1997).
[31] National Education Longitudinal Study. *National Education Longitudinal Study of 1988: Base Year Student Survey.* (Washington, DC: National Center for Education Statistics, 1995).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 2.3

Distribution of Science Scale Scores by Grade 4 Public School Students' Reports of Parents' Highest Education Level

	Average Scale Score	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
Did not finish HS						
DDESS	*** (****)	*** (****)	*** (****)	*** (****)	*** (****)	*** (****)
Nation	135 (2.2)	91 (3.2)	115 (3.4)	139 (3.0)	158 (2.8)	172 (3.7)
Graduated from HS						
DDESS	145 (2.4)	108 (10.5)	126 (3.1)	148 (3.4)	165 (3.5)	178 (8.2)
Nation	144 (1.6)	100 (2.9)	125 (1.5)	148 (2.7)	167 (1.3)	184 (2.1)
Some education after HS						
DDESS	156 (3.3)	120 (6.0)	135 (3.3)	158 (5.3)	177 (4.6)	194 (9.6)
Nation	154 (1.8)	110 (2.8)	136 (5.0)	157 (1.6)	176 (2.7)	192 (1.7)
Graduated from college						
DDESS	160 (1.7)	122 (5.1)	142 (2.9)	162 (2.2)	181 (1.7)	196 (2.9)
Nation	155 (1.3)	109 (3.4)	134 (1.9)	159 (1.4)	179 (1.2)	195 (2.3)
I don't know						
DDESS	149 (1.6)	113 (3.6)	131 (1.5)	151 (2.1)	170 (1.6)	185 (3.1)
Nation	142 (1.2)	100 (2.5)	122 (1.4)	145 (1.7)	164 (1.3)	182 (2.9)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Title I Participation

The Improving America's Schools Act of 1994 (P.L. 103-382) reauthorized the Elementary and Secondary Education Act of 1965 (ESEA). Title I Part A of the ESEA provides financial assistance to local educational agencies to meet the educational needs of children who are failing or most at risk of failing.[32] Title I programs are designed to help disadvantaged students meet challenging academic performance standards. Through Title I, schools are assisted in improving teaching and learning and in providing students with opportunities to acquire knowledge and skills outlined in their states' content and performance standards. For high poverty Title I schools, all children in the school may benefit through participation in schoolwide programs. Title I funding supports state and local education reform efforts and promotes coordination of resources to improve education for all students.

NAEP first collected student-level information on participation in Title I programs in 1994. The NAEP program will continue to monitor the performance of Title I program participants in future assessments. The Title I information collected by NAEP refers to current participation in Title I services. Students who participated in such services in the past but do not currently receive services are not identified as Title I participants. Differences between students who receive Title I services and those who do not should not be viewed as an evaluation of Title I programs. Typically, Title I services are intended for students who score poorly on assessments. To properly evaluate Title I programs, the performance of students participating in such programs must be monitored over time and their progress must be assessed.[33]

Table 2.4 presents results for fourth-grade students by Title I participation.

- The average scale score of DDESS students who were not receiving Title I services (154) was not significantly different from the national average (155).

[32] U.S. Department of Education, Office of Elementary and Secondary Compensatory Education Programs. *Improving Basic Programs Operated by Local Education Agencies.* (Washington, DC: U.S. Department of Education, 1996).

[33] For a study of mathematics performance of Title I students in 1991-1992, see U.S. Department of Education. *PROSPECTS: The Congressionally Mandated Study of Educational Growth and Opportunity, Interim Report: Language, Minority and Limited English Proficient Students.* (Washington, DC: U.S. Department of Education, 1995).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 2.4

Distribution of Science Scale Scores for Grade 4 Students by Title I Participation

	Average Scale Score	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
Participating						
DDESS	*** (****)	*** (****)	*** (****)	*** (****)	*** (****)	*** (****)
Nation	126 (2.0)	84 (2.4)	104 (1.9)	127 (2.9)	148 (1.8)	165 (3.5)
Not participating						
DDESS	154 (0.9)	116 (2.8)	134 (1.5)	155 (1.4)	175 (1.1)	190 (1.9)
Nation	155 (1.2)	115 (2.2)	137 (1.8)	158 (1.4)	177 (1.0)	192 (1.3)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

BEST COPY AVAILABLE

Free/Reduced-Price Lunch Program Eligibility

The free/reduced-price lunch component of the National School Lunch Program (NSLP), offered through the U.S. Department of Agriculture (USDA), is designed to ensure that children near or below the poverty line receive nourishing meals. [34] Eligibility for free or reduced prices for the meals is determined through the USDA's Income Eligibility Guidelines; it is included in this report as an indicator of poverty. The program is available to public schools, nonprofit private schools, and residential child care institutions.

NAEP first collected information on student-level eligibility for the federally funded NSLP in 1996. The NAEP program will continue to monitor the performance of these students in future assessments.

Table 2.5 shows the results for fourth graders based on their participation in this program.

- The average science scale score of students in DDESS who were eligible for free or reduced-price lunch (150) was higher than that of students nationwide (132). The average scale score of DDESS students who were not eligible for this service (160) was not significantly different from the national average (158).

- In DDESS, the average scale score of students who were eligible for free or reduced-price lunch (150) was lower than that of students who were not eligible (160). In the nation, eligible students also scored lower (132) than those who were not eligible (158).

[34] U.S. General Services Administration. *Catalog of Federal Domestic Assistance*. (Washington, DC: Executive Office of the President, Office of Management and Budget, 1995).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 2.5

Distribution of Science Scale Scores for Grade 4 School Students by Free/Reduced-Price Lunch Eligibility

	Average Scale Score	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
Eligible						
DDESS	150 (1.5)	115 (3.8)	132 (3.2)	151 (2.0)	168 (2.0)	184 (2.7)
Nation	132 (1.3)	88 (2.5)	110 (2.0)	134 (1.5)	156 (1.2)	174 (1.4)
Not eligible						
DDESS	160 (1.2)	122 (5.3)	142 (2.2)	162 (2.0)	180 (1.6)	195 (4.0)
Nation	158 (1.0)	121 (2.1)	140 (2.0)	160 (1.3)	178 (1.3)	193 (1.2)
Information not available						
DDESS	151 (1.9)	110 (6.3)	130 (1.6)	152 (2.1)	173 (2.9)	189 (4.2)
Nation	156 (6.0)	108 (10.5)	135 (6.3)	159 (5.3)	182 (7.4)	199 (3.9)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

PART TWO

Finding a Context for Understanding Students' Science Performance

The science performance of public school students in DDESS often can be better understood when viewed in the context of the environment in which the students are learning. This educational environment is largely determined by school characteristics, by characteristics of science instruction in the school, by home support for academics and other home influences, and by the students' own views about science. NAEP gathers information about this environment by means of the questionnaires administered to principals, teachers, and students.

Because NAEP is administered to a sample of students that is representative of the fourth-grade student population in the DDESS schools, NAEP results provide a view of the educational practices in DDESS that are useful for improving instruction and setting policy. However, despite the richness of the NAEP results, it is very important to note that NAEP data cannot establish a cause-and-effect relationship between educational environment and student scores on the NAEP science assessment.

The variables contained in Part Two are from the school characteristics and policies questionnaire, teacher questionnaires, and student background questionnaires. Part Two consists of four chapters: Chapter 3 discusses school characteristics related to science instruction;[35] Chapter 4 describes classroom practices related to science instruction, including curriculum, instructional emphasis, coursework, and computer use; Chapter 5 describes portions of a hands-on task and explores student exposure to these experiences; and Chapter 6 covers some potential influences from the home and from the students' own views about science.

To provide additional information, the bullets below sometimes contain results from one or more categories (i.e., from collapsed categories). When this is the case, the summed numbers reported in the bullets may be slightly different from the sums of the rounded numbers presented for each of the categories in the tables.

[35] Information on teacher preparation is included in Appendix E of this report.

CHAPTER 3

School Science Education Policies and Practices

School programs and conditions, instructional practices, and resource availability vary from state to state and even among schools within a locality. The information in this chapter is intended to give insight into those policies or practices that are associated with students' success in science.

The variables reported here reflect information from the questionnaires completed by principals and teachers of the public school students in the NAEP 1996 science assessment. In all cases, analyses are done at the student level. School and teacher-reported results are given in terms of the percentage of students who attend schools or who have teachers reporting particular practices.[36]

[36] Appendix A provides more details on the units of analysis used to derive the results presented in this report.

Emphasis on Science in the School

In the school characteristics and policies questionnaire, principals or other head administrators were asked several questions relating to the priority placed on science within their schools. Tables 3.1 and 3.2 present their responses.

- The percentage of fourth-grade students in DDESS attending schools that reported science was a priority (42 percent) was not different from the national percentage (42 percent). The average scale score for DDESS students in these schools (156) was significantly higher than that of students in schools nationwide reporting that science was a priority (146).

- The percentage of fourth-grade students who attended DDESS schools that reported having a district or state science curriculum that the school was expected to follow (87 percent) was not significantly different from the national percentage (92 percent).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 3.1

Schools' Reports on Science as a Priority at Grade 4

	Percentage and Average Scale Score	
	DDESS	**Nation**
Is this a school with a special focus on science?		
Yes	4 (0.1) *** (****)	4 (1.3) 153 (8.3)
Has your school identified science as a priority in the last two years?		
Yes	42 (0.8) 156 (1.7)	42 (4.7) 146 (1.9)
No	58 (0.8) 154 (1.3)	58 (4.7) 150 (1.3)
Does your district or state have a curriculum in science that your school is expected to follow?		
Yes	87 (0.5) 154 (1.1)	92 (2.3) 149 (0.9)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Principals were also asked how often students received science instruction. Schools using block scheduling (i.e., extended periods of instruction on fewer days) were not separately identified. Consequently, students in schools with block scheduling who received science instruction two or three times weekly may receive as many *hours* of instruction as students under traditional scheduling who receive instruction every day. Table 3.2 shows the following:

- The percentage of fourth-grade students in DDESS who attended schools that reported having instruction in science every day (68 percent) was higher than that of students across the nation (47 percent).

- The average scale score for DDESS students receiving science instruction every day (156) was higher than that of students nationwide receiving this much instruction (149).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 3.2

Schools' Reports on Time Spent in Science Instruction at Grade 4

How often does a typical fourth-grade student in your school receive instruction in science?	Percentage and Average Scale Score	
	DDESS	Nation
Twice a week or less/Not taught	7 (0.3) *** (****)	14 (3.3) 145 (3.3)
Three or four times a week	25 (0.5) 157 (2.1)	39 (4.4) 148 (2.4)
Every day	68 (0.6) 156 (1.3)	47 (4.2) 149 (2.3)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Resource Availability to Teachers

Resources available to teachers and schools vary. Past surveys have shown that teachers' perceptions of the availability of resources (e.g., materials, staff, and time) are variable across the country.[37] Previous NAEP assessments in other subject areas have shown an overall positive relationship in most states between teachers' reports of resource availability and their students' performance.[38]

[37] U.S. Department of Education. *Schools and Staffing in the United States: A Statistical Profile, 1993-94.* (Washington, DC: National Center for Education Statistics, 1996).

[38] For example, see Miller, K.E., J.E. Nelson, and M. Naifeh. *Cross-State Data Compendium for the NAEP 1994 Grade 4 Reading Assessment.* (Washington, DC: National Center for Education Statistics, 1995); National Center for Education Statistics. *State-by-State Background Questionnaire Data Appendix: NAEP 1992 Mathematics Assessment, Grades 4 and 8.* (Washington, DC: Office of Educational Research and Improvement, 1994).

Availability of Instructional Materials

Teachers often see the lack of resources and materials as a key problem for science instruction. In 1993 a national survey of elementary and secondary school educators reported that deficiencies related to instructional resources were the most serious problems for science instruction in their schools.[39] In that survey, schools reported spending a total of $0.51 per elementary student per year and $0.88 per middle grade student per year on science supplies, and $50 per year on science software (the average price for one piece of software is $100).

Teachers whose students participated in the NAEP 1996 science assessment were asked to categorize how well their school systems provided them with the classroom instructional materials they needed. The results are shown in Table 3.3.

- The percentage of students whose teachers reported receiving all of the resources they needed in DDESS (18 percent) was higher than that of students across the nation (10 percent).

- The average science scale score of students in DDESS whose teachers reported receiving all the resources they needed (159) was higher than that of the corresponding students in the nation (145).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 3.3

Teachers' Reports on Resource Availability at Grade 4

Which of the following statements is true about how well your school system provides you with the instructional materials and other resources you need to teach your class?	Percentage and Average Scale Score	
	DDESS	**Nation**
I get some or none of the resources I need.	13 (0.9) 150 (1.8)	41 (3.1) 147 (1.6)
I get most of the resources I need.	69 (1.4) 154 (1.3)	49 (3.1) 152 (1.3)
I get all the resources I need.	18 (1.1) 159 (1.4)	10 (1.7) 145 (2.7)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

[39] Weiss, I.R. *A Profile of Science and Mathematics Education in the United States: 1993.* (Chapel Hill, NC: Horizon Research, Inc., 1994).

Availability of Curriculum Specialist in the School

Table 3.4 shows the percentages and average scale scores of fourth-grade students in public school whose teachers indicated they had a curriculum specialist available to help or advise them in science.

- In DDESS, about half of the students were taught by teachers who reported having a curriculum specialist available to help or advise them in science (52 percent). This figure did not differ significantly from that of students across the nation (47 percent).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 3.4

Teachers' Reports on Curriculum Specialists at Grade 4

Is there a curriculum specialist available to help or advise you in science?	Percentage and Average Scale Score	
	DDESS	Nation
Yes	52 (1.4) 153 (1.4)	47 (3.6) 147 (1.5)
No	48 (1.4) 156 (1.4)	53 (3.6) 152 (1.6)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Parents as Classroom Aides

When school personnel and parents develop a positive line of communication, they strengthen the learning environment for the students both at school and at home. One of the most frequent reasons cited by school personnel for contacting parents is to request parent volunteer time at school.[40] The principals of the participating public schools were asked if parents were used as classroom aides. As shown in Table 3.5, principals for fourth graders reported the following.

- More than half of the students in DDESS (59 percent) were in schools that reported using parents as aides in classrooms routinely. In contrast, parents were not used as classroom aides for 6 percent of students in DDESS, according to school reports.

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 3.5

Schools' Reports on Parents as Aides in Classrooms at Grade 4

Does your school use parents as aides in classrooms?	Percentage and Average Scale Score	
	DDESS	Nation
No	6 (0.5) *** (****)	12 (2.7) 144 (4.2)
Yes, occasionally	35 (0.8) 155 (1.5)	46 (4.1) 148 (2.0)
Yes, routinely	59 (0.8) 154 (1.5)	42 (3.9) 150 (1.9)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
*** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

[40] U.S. Department of Education. *The Condition of Education 1995.* (Washington, DC: National Center of Education Statistics, 1995).

Student Absenteeism

School principals were asked if student absenteeism was a serious, moderate, minor problem, or not a problem. Table 3.6 shows results for fourth graders based on principals' reports.

- The percentage of students attending DDESS schools that reported that absenteeism was a moderate to serious problem (2 percent) was lower than that of students across the nation (15 percent).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 3.6

Schools' Reports on Student Absenteeism at Grade 4

To what degree is student absenteeism a problem in your school?	Percentage and Average Scale Score	
	DDESS	Nation
Not a problem	44 (0.9) 150 (1.6)	40 (4.1) 154 (2.3)
Minor	54 (0.9) 158 (1.4)	45 (4.3) 148 (2.0)
Moderate to serious	2 (0.3) *** (****)	15 (2.4) 136 (3.0)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

CHAPTER 4

Science Classroom Practices

Science education in the nation's schools has received considerable attention at the national, state, district, school, and classroom levels. In recent years, a number of national and international programs have measured student performance in science. The latest national trend report indicates that fourth graders have shown significant improvement compared to 1970.[41] A recent international study, the Third International Mathematics and Science Study (TIMSS), demonstrated that fourth-grade students' performance in the United States was above the international average compared to 26 countries;[42] students in only one country performed significantly higher.

Using guidance from such programs as the Statewide Systemic Initiative,[43] Project Scope, Sequence, and Coordination,[44] *Benchmarks for Science Literacy,*[45] and the *National Science Education Standards,*[46] many states are currently involved in evaluating their existing standards and developing new frameworks and criteria for science instruction in their state. TIMSS has also pointed out some differences between classroom practices in the United States and in the other participating nations that may guide development of more effective science instruction.[47]

This chapter focuses on curricular and instructional content issues in DDESS public schools and their relationship to students' science performance. For some of the issues discussed here, student- and teacher-reported results for similar questions are presented. In these situations, some discrepancies may exist between student- and teacher-reported percentages. It is not possible to offer conclusive reasons for these discrepancies or to determine which reports more accurately reflect fourth-grade classroom activities. The results presented give students' and teachers' impressions of the science classroom.

---

[41] Campbell, J.R., K. E. Voelkl, and P. L. Donahue. *NAEP 1996 Trends in Academic Progress.* (Washington, DC: National Center for Education Statistics, 1997).

[42] National Center for Education Statistics. *Pursuing Excellence: A Study of U. S. Fourth-Grade Mathematics and Science Achievement in International Context.* (Washington, DC: United States Government Printing Office, 1997).

[43] National Science Foundation, 1990 Statewide Systemic Initiative, provided grants to further research and initiatives in science reform.

[44] *Scope, Sequence and Coordination of Secondary School Science. Vol.1. The Content Core: A Guide for Curriculum Developers* (Washington, DC: National Science Teachers Association, 1992).

[45] American Association for the Advancement of Science, *Benchmarks for Science Literacy.* (New York: Oxford University Press, 1993).

[46] National Research Council. *National Science Education Standards.* (Washington, DC: National Academy Press, 1996).

[47] Martin, M. O., I.V.S. Mullis, A.E. Beaton, E.J. Gonzalez, T.A. Smith, and D.L. Kelly. *Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study.* (Boston: TIMSS International Study Center, 1997).

Curriculum Coverage

The NAEP 1996 science assessment examines three fields of science: earth, physical, and life. Fourth-grade public school teachers were asked how much time was spent on the three traditional fields of science in their classes and the results are presented in Table 4.1.

- In DDESS the percentage of fourth-grade public school students whose teachers reported spending a lot of time on earth science (32 percent) was significantly higher than that in the nation (18 percent). Students in DDESS classrooms where a lot of time was spent on earth science had an average scale score (155) that was not significantly different from that of students nationwide (148).

- The percentage of DDESS students whose teachers reported spending a lot of time on physical science (30 percent) was higher than the percentage nationwide (16 percent). The average science scale score in classrooms where physical sciences was covered a lot in DDESS (153) was not significantly different from the performance of students nationwide (152).

- The percentage of fourth-grade DDESS students whose teachers reported spending a lot of time on life science (40 percent) was significantly higher than the percentage nationwide (28 percent). The average scale score for students in these classrooms (156) was significantly higher than the average scale score across the nation (148).

THE NATION'S REPORT CARD naep
1996 State Assessment

TABLE 4.1

Teachers' Reports on Science Curriculum Coverage at Grade 4

How much time do you spend on each of the following areas of science in this class?		Percentage and Average Scale Score	
		DDESS	Nation
Earth science	None	1 (0.1) *** (****)	1 (0.3) *** (****)
	A little	3 (0.4) *** (****)	5 (1.1) 150 (4.3)
	Some	63 (1.1) 154 (1.1)	77 (2.7) 149 (1.1)
	A lot	32 (1.0) 155 (1.8)	18 (2.4) 148 (2.9)
Physical science	None	1 (0.1) *** (****)	2 (0.6) 137 (7.4)
	A little	7 (0.4) 153 (4.1)	9 (1.7) 144 (3.8)
	Some	61 (1.2) 155 (1.1)	73 (2.8) 149 (1.2)
	A lot	30 (1.2) 153 (2.0)	16 (2.5) 152 (3.0)
Life science	None	0 (****) *** (****)	1 (0.4) *** (****)
	A little	2 (0.3) *** (****)	6 (1.6) 149 (4.2)
	Some	58 (1.2) 153 (1.2)	65 (3.2) 150 (1.4)
	A lot	40 (1.1) 156 (1.7)	28 (3.1) 148 (1.8)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Fourth-Grade Students' Course Taking

Exposure to science and the opportunity to learn science have a positive effect on the science performance of students.[48] To investigate whether there is a relationship between science performance of students on the 1996 NAEP assessment and their study of science in school, information was collected relative to the amount of time spent each week on science instruction. As noted for Table 3.2, in which school principals answered a similar question concerning the frequency of science instruction, students with block scheduling were not identified separately. Based on students' responses shown in Table 4.2:

- In fourth grade, 2 percent of DDESS students reported never studying science in school. This is significantly lower than the nationwide percentage (4 percent).

- In DDESS schools, 43 percent of the students reported studying science every day, significantly higher than nationwide (31 percent). The average scale score for DDESS students who reported studying science every day (154) was significantly higher than that of students studying at this level nationwide (145).

[48] Council of Chief State School Officers. *State Indicators of Science and Mathematics Education.* (Washington, DC: CCSSO, 1995).

THE NATION'S REPORT CARD naep
1996 State Assessment

TABLE 4.2

Grade 4 Students' Reports on Their Science Classes

	Percentage and Average Scale Score	
	DDESS	Nation
About how often do you study science in school?		
Never	2 (0.4) *** (****)	4 (0.5) 131 (3.1)
Less than once a week	9 (0.8) 147 (3.3)	12 (1.1) 145 (2.3)
1 or 2 times a week	16 (1.0) 148 (3.8)	23 (1.2) 150 (1.0)
3 or 4 times a week	31 (1.4) 161 (1.6)	30 (1.4) 158 (1.5)
Every day	43 (1.2) 154 (1.4)	31 (1.9) 145 (1.6)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Instructional Emphasis

The framework that guided the development of the NAEP 1996 science assessment identified three ways of knowing and doing science — conceptual understanding, scientific investigation, and practical reasoning.[49] In addition, much focus in the science education reform effort has been placed on students' ability to communicate their understanding of science to others.[50] To assess students' opportunities to learn and communicate the knowledge and skills outlined in the framework, teachers were asked about their plans for science instruction during the entire year. Their responses are shown in Table 4.3

- In DDESS schools, 48 percent of the fourth-grade students had teachers who reported they planned to place moderate emphasis on the learning of science facts and terminology. This was not significantly different than the percentage of students nationwide whose teachers planned moderate emphasis on facts and terminology (56 percent).

- The average scale score of fourth-grade students whose teachers moderately emphasized science facts and terminology (154) was higher than that of their counterparts nationwide (149).

- In DDESS schools, 82 percent of the fourth-grade DDESS students had teachers who reported they planned to emphasize heavily the understanding of key science concepts by their students. Nationwide, a similar percentage of the students had teachers who planned heavy emphasis on conceptual understanding (78 percent).

- The average scale score of fourth-grade students whose teachers heavily emphasized understanding of key concepts was higher in DDESS schools (155) as compared to students in schools nationally (150).

- Teachers of 61 percent of the DDESS students reported they planned to emphasize heavily science problem-solving skills. Nationwide, the percentage of students was lower (49 percent).

- The average scale score of fourth-grade DDESS students with teachers placing heavy emphasis on problem-solving skills (156) was higher than that of students in the nation's public schools (150).

- In terms of learning how to communicate ideas in science effectively, 59 percent of the fourth-grade DDESS students had teachers who reported moderately emphasizing this ability for their students, and the percentage of comparable students nationwide (52 percent) was not significantly different.

---

[49] *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1995).

[50] American Association for the Advancement of Science. *Benchmarks for Science Literacy.* (New York: Oxford University Press, 1993); National Research Council. *National Science Education Standards.* (Washington, DC: National Academy Press, 1996).

- The average scale score of fourth-grade students whose teachers placed moderate emphasis on communicating science ideas was significantly higher in DDESS schools (153) as compared to schools in the national sample (148).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 4.3

Teachers' Reports on Instructional Emphasis at Grade 4

Think about your plans for your science instruction during the entire year. About how much emphasis will you give to the following as an objective for your students?	**Percentage and Average Scale Score**	
	DDESS	**Nation**
Knowing science facts and terminology		
Little or no emphasis	3 (0.4) *** (****)	3 (1.1) 158 (6.7)
Moderate emphasis	48 (1.2) 154 (1.4)	56 (3.2) 149 (1.5)
Heavy emphasis	49 (1.2) 156 (1.4)	41 (2.9) 148 (1.7)
Understanding key science concepts		
Little or no emphasis	0 (****) *** (****)	0 (****) *** (****)
Moderate emphasis	18 (1.1) 151 (2.1)	22 (2.1) 145 (2.1)
Heavy emphasis	82 (1.1) 155 (1.1)	78 (2.1) 150 (1.0)
Developing science problem-solving skills		
Little or no emphasis	3 (0.8) *** (****)	6 (1.7) 158 (4.1)
Moderate emphasis	36 (1.3) 153 (1.6)	45 (3.1) 147 (1.6)
Heavy emphasis	61 (1.3) 156 (1.2)	49 (3.3) 150 (1.4)
Knowing how to communicate ideas in science effectively		
Little or no emphasis	7 (0.9) 151 (4.3)	12 (2.1) 154 (4.1)
Moderate emphasis	59 (1.5) 153 (1.2)	52 (3.5) 148 (1.3)
Heavy emphasis	34 (1.3) 157 (1.6)	35 (3.8) 150 (1.7)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

With the explosion of the information age, mainstream news and the Internet afford opportunities to access up-to-date scientific information. Science instruction could benefit by taking advantage of such opportunities. To determine if these opportunities were being explored, fourth-grade teachers and students were asked how often they have classroom discussions about science stories that appear in the news. The results are presented in Table 4.4; some categories in the table have been combined for the bullets below.

- In DDESS schools, 39 percent of fourth-grade students were taught by teachers who reported once- or twice-weekly classroom discussions of science in the news. This was significantly higher than the percentage nationwide (31). The average scale score of these DDESS students (154) was not significantly different from that of students nationwide whose teachers reported discussions of science in the news this often. (149).

- When students were asked how often they discussed science in the news, 16 percent in DDESS schools reported once- or twice-weekly discussions, while 15 percent of the nation's public school students reported discussions of science in the news this often.

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 4.4

Teachers' and Students' Reports on Discussions of Science in the News

How often do your students (do you) discuss science in the news?	Percentage and Average Scale Score			
	DDESS		Nation	
	Teacher	Student	Teacher	Student
Never or hardly ever	13 (0.9) 153 (2.6)	60 (1.7) 156 (1.2)	20 (2.7) 149 (2.4)	58 (1.1) 152 (0.9)
Once or twice a month	43 (1.3) 154 (1.3)	12 (1.1) 156 (3.5)	46 (3.5) 150 (1.8)	15 (0.9) 154 (1.6)
Once or twice a week	39 (1.1) 154 (1.6)	16 (1.2) 152 (2.4)	31 (3.0) 149 (1.9)	15 (0.7) 147 (1.6)
Almost every day	5 (0.5) *** (****)	13 (0.9) 145 (3.9)	4 (1.5) 137 (12.4)	12 (0.8) 135 (2.0)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Science Homework

Past NAEP science assessments have shown a positive relationship between science homework and performance.[51] To examine the relationship between homework and science scale scores in DDESS schools, the teachers of the assessed students were asked to report the amount of science homework they assigned each week, and students were asked to report the amount of time they spent on science homework each week.

Tables 4.5 and 4.6 show the fourth-grade science teachers' and students' responses. Since students had an additional response choice, "I am not taking a science course this year," but no analogous option was available to teachers, the results are reported in separate tables. According to the teachers' responses:

- In DDESS schools, teachers reported 31 percent of the fourth graders were assigned a half hour of science homework each week. Public school teachers nationally reported assigning this same amount of homework to a percentage of students that was not significantly different (39 percent). For such students, DDESS fourth graders' scores were significantly higher (155) than those of fourth graders in the nation's public schools (148).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 4.5

Teachers' Reports on Homework in Science at Grade 4

About how much time do you expect a student in this class to spend doing homework each week?	Percentage and Average Scale Score	
	DDESS	Nation
None	16 (1.2) 149 (1.4)	22 (2.6) 152 (2.7)
1/2 hour	31 (1.2) 155 (1.9)	39 (3.5) 148 (2.0)
1 hour	42 (1.4) 154 (1.4)	31 (3.2) 149 (2.2)
2 hours	9 (1.1) 159 (3.1)	6 (1.3) 147 (7.1)
More than 2 hours	1 (0.1) *** (****)	2 (0.7) 141 (7.8)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

[51] Jones, L.R., I.V.S. Mullis, S.A.Raizen, I.R. Weiss, and E.A. Weston. *The 1990 Science Report Card: NAEP's Assessment of Fourth, Eighth, and Twelfth Graders.* (Washington, DC: National Center for Education Statistics, 1992).

The fourth-grade students' reports indicated that:

- For fourth graders reporting spending no time on science homework in a typical week, the percentage for DDESS schools (23 percent) and the nations public schools (25 percent) did not differ significantly, and the average scale score for DDESS students (153) did not differ significantly from that for similar students nationally (152).

THE NATION'S REPORT CARD NAEP 1996 State Assessment

TABLE 4.6

Grade 4 Students' Reports on Homework in Science

About how much time do you spend doing science homework each week?	Percentage and Average Scale Score	
	DDESS	Nation
I don't have science.	10 (1.2) 143 (3.4)	13 (1.0) 144 (1.9)
None	23 (1.2) 153 (2.1)	25 (1.2) 152 (1.3)
1/2 hour	45 (1.4) 159 (1.4)	39 (1.2) 153 (1.2)
1 hour	14 (1.3) 151 (3.7)	15 (0.8) 146 (1.9)
2 hours	4 (0.6) *** (****)	3 (0.3) 140 (3.7)
More than 2 hours	3 (0.4) *** (****)	4 (0.4) 130 (2.5)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

In addition to being asked about science homework in general, students were asked how often they use a computer at home for schoolwork. The question was not restricted to science homework, so students' reports most likely included homework for other academic areas such as English and mathematics. Given the trend that home computers are steadily assuming more importance for completing homework assignments,[52] it seems useful that NAEP monitor the prevalence of this practice and its relationship to performance.

Based on the reports of fourth graders in DDESS, as shown in Table 4.7:

- For DDESS students, 38 percent had no computer at home. This was not significantly different than the percentage for the nation's students (43 percent). However, there was a significant difference in average scale scores for these DDESS students (150) and the students in the nation (143).

- Of the fourth graders who reported using their home computer to do homework almost every day, the percentage of DDESS students (10 percent) did not differ significantly from the percentage of students in the nation (11 percent). The average scale score for these DDESS students (143) was not significantly higher than that for the nation's students who used their home computers for homework almost daily (138).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 4.7

Grade 4 Students' Reports on Using Computers at Home

How often do you use a computer at home for schoolwork?	Percentage and Average Scale Score	
	DDESS	Nation
There is no computer at home	38 (1.5) 150 (1.2)	43 (1.7) 143 (1.1)
Never or hardly ever	34 (1.4) 156 (1.8)	25 (0.9) 155 (1.3)
Once or twice a month	10 (1.0) 164 (3.4)	10 (0.7) 161 (1.5)
Once or twice a week	8 (0.6) 161 (3.1)	10 (0.7) 154 (2.0)
Almost every day	10 (0.7) 143 (3.1)	11 (0.6) 138 (2.0)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

[52] National Center for Education Statistics. *Digest of Education Statistics 1995*. (Washington, DC: National Center for Education Statistics, 1995).

Computer Use in Science Instruction

The use of computers in the collection of data, interpretation of results, and communication of findings is part of the *Benchmarks for Science Literacy* and the recently published *National Science Education Standards*.[53] Recommendations for facilitating science instruction in the nation's schools often include more use of computers. Computers can be used to demonstrate scientific concepts, simulate scientific phenomena, deliver instruction, and collect and analyze data. Of course, effective computer use may depend on many factors other than availability, such as teachers' training or whether computers have been incorporated into the curriculum effectively.

Given the potential role of computers in science instruction, NAEP asked DDESS students and their teachers about the availability and use of computers in science instruction. As presented in Table 4.8, when fourth-grade DDESS science teachers were asked about the availability of computers, their responses indicated the following:

- In DDESS schools, 9 percent of the students had teachers who reported that no computers were available for use in their science classes; this was not significantly different than at the national level (14 percent). The average scale scores for these students of DDESS and national public school teachers (149 and 141, respectively) were not significantly different.

- In DDESS schools, the percentage of teachers reporting that their students had access to four or more computers in the classroom (30 percent) was higher than that for the nation (10 percent). The average scale score of DDESS students whose teachers reported four or more computers in the classroom (153) was not significantly different than that of students in the national sample (152).

[53] American Association for the Advancement of Science. *Benchmarks for Science Literacy.* (New York: Oxford University Press, 1993); National Research Council. *National Science Education Standards* (Washington, DC: National Academy Press, 1996).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 4.8

Teachers' Reports on the Availability of Computers at Grade 4

Which best describes the availability of computers for use by your science students?	Percentage and Average Scale Score	
	DDESS	Nation
None available	9 (0.6) 149 (1.6)	14 (2.0) 141 (3.8)
One within the classroom	20 (1.0) 154 (2.6)	27 (4.0) 147 (2.5)
Two or three within the classroom	18 (1.0) 157 (1.6)	18 (2.5) 148 (2.8)
Four or more within the classroom	30 (1.0) 153 (2.2)	10 (2.6) 152 (4.9)
Available in a computer laboratory but difficult to access or schedule	12 (0.8) 158 (2.7)	13 (2.9) 159 (3.4)
Available in a computer laboratory and easy to access or schedule	11 (1.0) 158 (3.0)	18 (3.2) 147 (2.9)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

The availability of computers varies from school to school and the uses for computers can vary widely from class to class. Computers can be used in many ways to help students learn science, including simulating scientific phenomena or illustrating models. Also, the frequency of use can vary, regardless of the primary use in the classroom. Teachers in DDESS schools were asked how they used computers and how often they were used in their science classroom. Also, students were asked how often they used computers when doing science in school. The responses of fourth-grade teachers to the purpose of use for science instruction, as shown in Table 4.9, indicate the following:

- The percentage of DDESS students whose teachers reported using computers instructionally with science or learning games (40 percent) was higher than the corresponding national percentage (29 percent). The average scale score for these DDESS students (154) was not significantly different from students in the nation (151).

- The percentage of DDESS students whose teachers reported that they did not use computers for instruction in science (40 percent) was lower than the percentage of comparable students nationwide (52 percent). The average scale score of these DDESS students (154) was higher than that of students in the national sample (146).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 4.9

Teachers' Reports on the Use of Computers for Instruction in Science at Grade 4

How do you use computers for instruction in science?	Percentage and Average Scale Score	
	DDESS	**Nation**
Drill and practice	8 (0.8) 153 (3.4)	5 (1.6) 145 (6.4)
Playing science/learning games	40 (1.3) 154 (1.8)	29 (2.9) 151 (2.0)
Simulations and modeling	21 (1.3) 152 (2.6)	19 (3.1) 153 (2.1)
Data analysis and other applications	4 (0.3) *** (****)	6 (1.4) 147 (5.2)
Word processing	12 (0.7) 152 (3.0)	10 (1.8) 157 (3.2)
I do not use computers for science instruction	40 (1.3) 154 (1.5)	52 (3.2) 146 (1.6)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

BEST COPY AVAILABLE

Table 4.10 presents teacher and student reports on the frequency of use of computers for science.

- Significantly fewer DDESS students (56 percent) than students in the nation's public schools (69 percent) had teachers who reported that they never or hardly ever used a computer for science instruction. The average scale score for students of these DDESS teachers (154) was higher than the scale score for the students of such teachers nationwide (148).

- In DDESS, 69 percent of the students reported never or hardly ever using computers to do science in school. This was not significantly different from the percentage of students at the national level (67 percent). These two groups had average scale scores that were not significantly different (157 for DoDDS, 153 for the nation).

- The percentages of students using computers for science almost every day in DDESS schools (10 percent) and national public schools (10 percent) were the same; however, the average score for the DDESS students using computers almost every day (143) was higher than that of students in the national sample (130).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 4.10

Teachers' and Students' Reports on the Frequency of Computer Use at Grade 4

How often do your students (do you) use a computer for science?	Percentage and Average Scale Score			
	DDESS		Nation	
	Teacher	Student	Teacher	Student
Never or hardly ever	56 (1.4) 154 (1.1)	69 (1.4) 157 (1.1)	69 (4.0) 148 (1.7)	67 (1.4) 153 (0.9)
Once or twice a month	26 (1.5) 157 (2.0)	12 (1.0) 158 (4.2)	20 (2.9) 153 (1.8)	12 (0.7) 152 (2.1)
Once or twice a week	14 (0.8) 152 (2.3)	10 (1.0) 145 (4.6)	10 (2.4) 148 (3.5)	11 (0.8) 147 (2.0)
Almost every day	5 (0.2) *** (****)	10 (1.1) 143 (3.5)	2 (0.7) 150 (8.6)	10 (0.5) 130 (1.9)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

CHAPTER 5

Student Performance on Hands-On Science Tasks

A number of goals for science education have been put forward in a series of reports authored by government agencies and professional societies over the last 15 years.[54] These goals include acquisition of a core of scientific understanding, ability to apply science knowledge in practical ways, familiarity with experimental design, and the ability to carry out scientific experiments. The reports also offered recommendations for the science curricula and instruction needed to achieve these goals. One recommendation was to encourage active student participation in hands-on science, learning in cooperative groups, and completing sustained projects.[55]

A 1993 national survey indicated that fourth-grade science teachers devote as much as 26 percent of class time to hands-on, or manipulative, activities.[56] NAEP included assessments of higher-order thinking skills in science and mathematics as early as 1986, through a pilot assessment that required students to work on various hands-on tasks. Although the NAEP 1990 science assessment measured skills that were integral to scientific investigation,[57] hands-on tasks were not included. When the 1996 science framework[58] was developed in the early 1990s, it took into account the current reforms in science education by specifying three question types that probed understanding of conceptual and reasoning skills: performance exercises, constructed-response questions, and multiple-choice questions. It was envisaged that in the performance exercises, students would manipulate selected physical objects and try to solve a scientific problem using the objects before them. Hands-on tasks that met these criteria were developed for the 1996 science assessment, and each student who participated in the assessment was given an opportunity to conduct one of them.

[54] National Science Board Commission on Precollege Education in Mathematics, Science, and Technology. *Educating America for the 21st Century.* (Washington, DC: National Science Foundation, 1983); American Association for the Advancement of Science. *Science For All Americans: A Project 2061 Report on Literacy Goals in Science, Mathematics, and Technology.* (Washington, DC: American Association for the Advancement of Science, 1989); Aldridge, B.G. *Essential Changes in Secondary School Science: Scope, Sequence, and Coordination.* (Washington, DC: National Science Teachers Association, 1989); National Research Council. *Fulfilling the Promise: Biology Education in the Nation's Schools.* (Washington, DC: National Academy Press, 1990).

[55] *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1993).

[56] Rolf K. Blank and Doreen Gruebel. *State Indicators of Science and Mathematics Education, 1995.* (Washington, DC: Council of Chief State School Officers, 1995).
In the TIMSS, teachers report spending 25% of class time on hands-on activities. Schmidt, W.H., et al. *TIMSS Results: Curriculum, Instruction, and Achievement* AAAS Annual Meeting, Seattle, WA, February 14, 1997.

[57] *Science Objectives: 1990 Assessment.* (Princeton, NJ: The National Assessment of Educational Progress, 1989).

[58] *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1995).

NAEP Hands-On Science Tasks

Four different hands-on tasks were administered in the NAEP 1996 science assessment. Each task was designed to use materials to perform investigations, make observations, evaluate experimental results, and apply problem-solving skills. In addition, tasks shared the following characteristics.

- Diagrams were included to guide students through the procedures.
- Multiple-choice and constructed-response questions were embedded throughout the task.
- Scientific investigation was integrated with conceptual understanding and practical reasoning.

The creation of the hands-on tasks presented special challenges. Since the assessment was administered in a variety of settings ranging from laboratories to cafeterias, all of the required equipment necessary to conduct each task had to be provided in a self-contained kit produced according to standard specifications to ensure uniformity. There were some limitations on materials and equipment. For example, live materials (with the exception of seeds) and equipment that required an electric outlet were not used. Safety was also an important concern and was addressed in a number of ways. Each state's safety regulations were considered; no toxic or corrosive chemicals were used; assessment administrators were trained in appropriate laboratory safety; and students were provided with goggles for some tasks.

Sample Questions from a Task

A brief summary of one of the four tasks given to grade 4 students in DDESS appears below. In Figure 5.1, the materials for the task are described. Two sample questions with a student response appear in Figures 5.2 and 5.3. (Note: the student responses and the percentages of students receiving complete or partial scores are from the national sample, and do not necessarily reflect performance of students in the DoDEA schools).

THE NATION'S REPORT CARD naep 1996 State Assessment

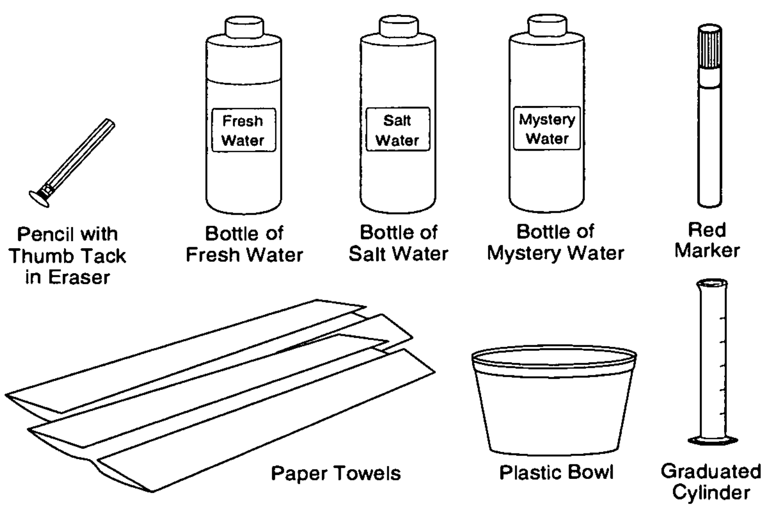
FIGURE 5.1

Materials for the Grade 4 Hands-On Task: Floating Pencil

FLOATING PENCIL

Using a Pencil to Test Fresh and Salt Water

You have been given a bag with some things in it that you will work with during the next 20 minutes. Take all of the things out of the bag and put them on your desk. Now look at the picture below. Do you have everything that is shown in the picture? If you are missing anything, raise your hand and you will be given the things you need.



An instrument constructed from a pencil and thumbtack served as a hydrometer in this task. Students were asked to observe, measure, and compare the lengths of a portion of pencil, marked with calibrations for ease of measurement, that floated above the water surface in fresh water and salt water. The students then determined if an unknown water sample was fresh water or salt water and predicted how the addition of more salt to the salt water would affect the floating pencil. The task assessed students' ability to make simple observations, measure volume using a graduated cylinder, measure length using a ruler, apply observations and measurements to test an unknown, make generalized inferences from observations, and apply understanding to an everyday situation.

Figure 5.2 presents a short constructed-response question that asks students to use the floating-pencil test to find out if the water in a bottle labeled "Mystery Water" is fresh water or salt water and explain how they are able to tell. This question was presented towards the end of the task after students had measured the height of the pencil above the fresh water, salt water, and the mystery water. Responses to this question were scored according to a three-level guide: *Complete, Partial,* or *Incorrect*. Figure 5.2 also presents a sample of a student response that received a score of *Complete*. The response received a score of *Complete* because the mystery water was identified and the explanation specifically referred to the level the fresh water and the mystery water reached on the calibrated pencil. Twenty-eight percent of students were able to correctly identify the mystery water and give a satisfactory explanation.

THE NATION'S REPORT CARD naep 1996 State Assessment

FIGURE 5.2

Sample Question from the Grade 4 Hands-On Task: Floating Pencil

Students' responses were scored using a three-level scoring guide that allowed for partial credit. The sample student response received the highest score, **Complete**, because it stated that the mystery water was fresh water and gave a satisfactory explanation that referred to observations made when doing the hands-on task.

Percentages of Fourth Graders Receiving Complete and Partial Scores	
Complete	**28%**
Partial	**45%**

Is the mystery water fresh water or is it salt water?

fresh water

How can you tell what the mystery water is?

because in the fresh water it went to A and it went to A again

Figure 5.3 presents a short constructed-response question that asks students to apply their observations of the behavior of a pencil in different solutions to a real-world situation (swimming in salt water and fresh water). This question was presented at the end of the task after students had measured the height of the pencil above the fresh water, salt water, and the mystery water and determined what the mystery water was. Responses to this question were scored according to a three-level guide: *Complete, Partial,* or *Incorrect.* Figure 5.3 also presents a sample of a student response that received a score of *Complete*. The ocean was correctly identified and the explanation referred to information learned by performing the hands-on task. Fourteen percent of students were able to apply their findings.

THE NATION'S REPORT CARD naep 1996 State Assessment

FIGURE 5.3

Sample Question from the Grade 4 Hands-On Task: Floating Pencil

Students' responses were scored using a three-level scoring guide that allowed for partial credit. The sample student response received the highest score, **Complete**, because it stated that it was easier to stay afloat in the ocean and gave a satisfactory explanation that referred to information learned while conducting the hands-on task.

Percentages of Fourth Graders Receiving Complete and Partial Scores	
Complete	**14%**
Partial	**29%**

When people are swimming, is it easier for them to stay afloat in the ocean or in a freshwater lake?

ocean

Explain your answer.

The ocean has salt water and the pencail stayed higher a'float in salt water than fresh water.

BEST COPY AVAILABLE

Instruction Related to Scientific Investigation

Fourth-grade science teachers at DDESS schools were asked about the emphasis they placed on laboratory skills and data analysis in their science classes and about the frequency and nature of hands-on activities or investigations assigned by them. Students were asked about the frequency and nature of hands-on activities or investigations conducted by them.

As mentioned before, a direct cause-and-effect relationship between educational environment and student scores on the NAEP science assessment is not implied. However, responses to teacher and school questionnaires provide a broad view of educational practices that should prove useful for improving instruction and setting policy. The teachers' and students' responses are presented in Tables 5.1 through 5.5.

- The percentage of fourth-grade students in DDESS schools whose teachers reported placing moderate emphasis on the development of laboratory skills and techniques (42 percent) was smaller than the percentage nationwide (56 percent). Students whose teachers reported moderate emphasis on laboratory skills and techniques in DDESS had an average scale score (157) which was higher than that of students nationwide (148).

- The percentage of fourth-grade DDESS students whose teachers reported heavy emphasis on the development of data analysis skills (18 percent) was significantly higher than that of students nationwide (12 percent). Fourth-grade students whose teachers reported heavy emphasis on data analysis skills had an average science scale score (155) which was not significantly different from that of students whose teachers reported heavy emphasis on the development of data analysis skills (147).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 5.1

Grade 4 Teachers' Reports on Science Instruction Related to Performance Tasks

Think about your plans for your science instruction during the entire year. About how much emphasis will you give to each of the following?	Percentage and Average Scale Score	
	DDESS	Nation
Developing laboratory skills and techniques as an objective for your students		
Little or no emphasis	28 (1.1) 150 (1.9)	29 (2.7) 149 (1.7)
Moderate emphasis	42 (1.6) 157 (1.2)	56 (2.7) 148 (1.3)
Heavy emphasis	30 (1.4) 156 (1.9)	14 (2.0) 153 (3.0)
Developing data analysis skills		
Little or no emphasis	16 (1.1) 151 (3.2)	35 (3.0) 149 (2.2)
Moderate emphasis	66 (1.4) 155 (1.2)	53 (3.2) 150 (1.4)
Heavy emphasis	18 (1.0) 155 (2.3)	12 (1.9) 147 (3.9)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

BEST COPY AVAILABLE

The percentages of students exposed to classroom science demonstrations with a given frequency may vary — depending on whether reported by the teacher or the students. However, it is not possible to determine reasons for these discrepancies, although it is probably true that perceptions of teachers and their students may sometimes differ greatly.

- Teachers who reported doing a science demonstration once or twice a month taught 36 percent of DDESS fourth-grade students, which was not significantly different from the percentage of students nationwide (44 percent) whose teachers did science demonstrations with the same frequency. However, these DDESS students' average scale scores (153) were not significantly different from those of their counterparts in the nation's public schools (148).

- The percentage of fourth-grade DDESS students (23 percent) reporting that their teachers did science demonstrations once or twice a month did not differ significantly from the percentage of such students nationally (27 percent). The DDESS students had an average scale score (161) that did not differ significantly from that of their national counterparts (158).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 5.2

Teachers' and Students' Reports on the Frequency of Science Demonstrations at Grade 4

How often do you (does your teacher) do a science demonstration?	Percentage and Average Scale Score			
	DDESS		Nation	
	Teacher	Student	Teacher	Student
Never or hardly ever	7 (0.4) 148 (4.5)	37 (1.6) 153 (1.5)	7 (1.5) 153 (2.7)	41 (1.5) 150 (1.1)
Once or twice a month	36 (1.2) 153 (1.5)	23 (1.6) 161 (2.5)	44 (4.1) 148 (1.5)	27 (0.7) 158 (1.3)
Once or twice a week	42 (1.4) 157 (1.6)	23 (1.2) 156 (2.1)	46 (4.2) 149 (2.1)	22 (1.2) 148 (1.8)
Almost every day	16 (0.9) 154 (2.0)	16 (1.1) 149 (2.6)	4 (1.1) 155 (8.4)	10 (0.7) 136 (2.3)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

- The percentage of fourth-grade DDESS students whose teachers reported that their science students performed hands-on tasks once or twice a week (48 percent) was not significantly different from the nationwide percentage (47 percent).

- The percentage of fourth-grade DDESS students whose teachers reported that their students did hands-on tasks once or twice a week had an average science scale score (157) which was significantly higher than that of students nationwide whose teachers reported this same level of hands-on task experience (150).

- The percentage of DDESS students reporting that they do hands-on projects once or twice a week (26 percent) is not different from that for the nation's fourth graders (26 percent). The average scale score for DDESS students reporting the same frequency of hands-on activity (155) is not significantly higher than that for the nation (152).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 5.3

Teachers' and Students' Reports on the Frequency of Hands-on Activities or Investigations at Grade 4

How often do your students (do you) do hands-on activities or investigations in science?	Percentage and Average Scale Score			
	DDESS		Nation	
	Teacher	Student	Teacher	Student
Never or hardly ever	3 (0.5) *** (****)	23 (1.3) 154 (2.1)	3 (1.1) 142 (5.2)	28 (1.4) 149 (1.2)
Once or twice a month	34 (1.3) 154 (1.5)	25 (1.3) 162 (2.1)	41 (3.5) 149 (1.8)	27 (1.1) 158 (0.9)
Once or twice a week	48 (1.1) 157 (1.5)	26 (1.2) 155 (1.8)	47 (3.2) 150 (1.5)	26 (1.2) 152 (1.8)
Almost every day	15 (0.7) 148 (1.7)	26 (1.1) 146 (2.2)	9 (1.8) 146 (3.4)	19 (0.9) 138 (2.0)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

BEST COPY AVAILABLE

- The same percentage of fourth-grade students in DDESS schools and in the nation (75 percent) had teachers who reported assigning science projects in school which take a week or more to complete. The average scale score for these DDESS students (156) was higher than that for students in the nation's public schools (150).

- The same percentages of fourth-grade students in DDESS and in the nation who reported doing science projects or investigations that take a week or more were the same (60 percent), which was higher than percentages of DDESS students who did not (40 percent). The average scale score of these DDESS students (154) was higher than that of students in the nation's public schools (148).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 5.4

Teachers' and Students' Reports on Long-Term Science Projects at Grade 4

Do you ever assign (do) individual or group science projects or investigations in school that take a week or more?	Percentage and Average Scale Score			
	DDESS		Nation	
	Teacher	Student	Teacher	Student
Yes	75 (1.3) 156 (1.0)	60 (1.4) 154 (1.3)	75 (3.1) 150 (1.1)	60 (1.5) 148 (1.1)
No	25 (1.3) 151 (2.4)	40 (1.4) 154 (1.5)	25 (3.1) 146 (2.2)	40 (1.5) 149 (1.2)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

CHAPTER 6

Influences Beyond School that Facilitate Learning Science

The home environment can be an important support for the school environment. To examine the relationship between science scale scores and home factors, data regarding students' responses to questions about home factors and principals' responses to questions about parental involvement in the school were examined. In order to examine the impact of student mobility on academic achievement, the student questionnaires also asked students how often they had changed schools because of household moves.

Students' attitudes towards science probably influence their performance in the assessment. Their attitudes towards science may be attributed to factors within the school as well as to external influences. In the recent TIMSS survey, for fourth grade students in more than one-third of the countries, a positive relationship existed between liking science and science achievement. Although the pattern was not uniform across countries, the students who reported liking science or liking it a lot had higher achievement than those who reported disliking it to some degree.[59]

[59] Martin, M. O., I.V.S. Mullis, A.E. Beaton, E.J. Gonzalez, T.A. Smith, and D.L. Kelly. *Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study.* (Boston: TIMSS International Study Center, 1997).

Discussing Studies at Home

The importance of schoolwork for students and their families can be measured by how often it is discussed at home. When students discuss academic work at home, they create an important link between home and school. Recent NAEP assessments in various subject areas have found a positive relationship between discussing studies at home and student performance.[60]

The NAEP 1996 assessment asked students to report on how frequently they discuss schoolwork at home. As shown in Table 6.1, the results for fourth graders attending DDESS schools indicate that:

- The percentage of students who said they discussed schoolwork with someone at home once or twice a week was not significantly different in DDESS schools (18 percent) than in the nation's public schools (21 percent). The average scale scores for these two groups (153 for DDESS, 151 for the nation) were not significantly different.

- The average scale score for DDESS students who discussed their schoolwork almost every day (156) was higher than that for the nation's students (150); however, the percentages of students in this category in DDESS and the nation did not differ significantly (54 and 53, respectively).

[60] Campbell, J.R., P.L. Donahue, C.M. Reese, and G.W. Phillips. *NAEP 1994 Reading Report Card for the Nation and the States.* (Washington, DC: National Center for Education Statistics, 1996); Beatty,A.S., C.M. Reese, H.R. Persky, and P. Carr. *NAEP 1994 U.S. History Report Card.* (Washington, DC: National Center for Education Statistics, 1996); Persky, H.R., C.M. Reese, C.Y. O'Sullivan, S. Lazer, J. Moore, and S. Shakrani. *NAEP 1994 Geography Report Card.* (Washington, DC: National Center for Education Statistics, 1996).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 6.1

Grade 4 Students' Reports on Discussing Studies at Home

How often do you discuss things you have studied in school with someone at home?	Percentage and Average Scale Score	
	DDESS	Nation
Never or hardly ever	20 (1.2) 148 (2.4)	19 (0.9) 142 (1.6)
Once or twice a month	8 (0.9) 152 (3.8)	7 (0.4) 143 (2.3)
Once or twice a week	18 (1.0) 153 (2.6)	21 (0.7) 151 (1.4)
Almost every day	54 (1.6) 156 (1.1)	53 (1.1) 150 (1.0)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Literacy Materials in the Home

Students can learn much about science by reading materials outside the classroom. For example, scientific information can often be found in mainstream newspaper and magazine articles. Also, the availability of reading and reference materials at home may be an indicator of the value placed on learning by the parents.[1] TIMSS reported that in most countries, the more books students reported in their homes, the higher their science achievement.[2] In recent NAEP assessments, a positive relationship has been reported between print materials in the home and average scale scores.[3]

The NAEP assessment asked students whether their families used more than 25 books, an encyclopedia, a newspaper, or any magazines in their home. Table 6.2 shows the percentages of fourth-grade public school students reporting that their families have all four types, only three types, or two or fewer types of these literacy materials and the corresponding students' average scale scores. Based on their responses:

- About one third of the DDESS students (34 percent) reported having all four types of literacy materials in their homes. This percentage was not significantly different from the percentage for the nation (36 percent).
- In comparison, the percentage of DDESS students reporting having two or fewer types of these materials (29 percent) was smaller than the percentage having all four types (34 percent). For the nation, the percentage having two or fewer types (33 percent) was not significantly different from the percentage having all four types (36 percent).
- The average science scale score for DDESS students with all four types of literacy materials (162) was greater than that for students with two or fewer types (147).

[1] Rogoff, B., *Apprenticeship in Thinking: Cognitive Development in Social Context.* (New York: Oxford University Press, 1990).
[2] Martin, M. O., I.V.S. Mullis, A.E. Beaton, E.J. Gonzalez, T.A. Smith, and D.L. Kelly. *Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study.* (Boston: TIMSS International Study Center, 1997).
[3] Campbell, J.R., P.L. Donahue, C.M. Reese, and G.W. Phillips. *NAEP 1994 Reading Report Card for the Nation and the States.* (Washington, DC: National Center for Education Statistics, 1996); Beatty, A.S., C.M. Reese, H.R. Persky, and P. Carr. *NAEP 1994 U.S. History Report Card.* (Washington, DC: National Center for Education Statistics, 1996).

THE NATION'S REPORT CARD naep
1996 State Assessment

TABLE 6.2

Grade 4 Students' Reports on Literacy Materials in the Home

How many of the following types of reading materials are in your home (more than 25 books, an encyclopedia, a newspaper, magazines)?	Percentage and Average Scale Score	
	DDESS	Nation
Zero to two	29 (1.2) 147 (1.8)	33 (1.2) 137 (1.3)
Three	37 (1.2) 151 (1.8)	31 (0.6) 150 (1.0)
Four	34 (1.3) 162 (1.6)	36 (1.4) 157 (1.2)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Television Viewing Habits

The recent TIMSS report discusses television watching, and compares it with amount of time spent on other activities, including homework. It was found that the relationship between science achievement and amount of time spent watching television was similar to the relationship between achievement and time spent on homework. Watching less than one hour per day was associated with lower academic achievement; perhaps low television watching is a surrogate socioeconomic indicator. Watching from one to two hours per day was associated with the highest science achievement.[64]

Past NAEP assessments have shown that over 40 percent of fourth-grade students reported watching four or more hours of television each day. A major concern is that time spent watching television reduces the time spent on homework and related academic activities. Although the effects of such extensive television exposure are difficult to document, a generally negative relationship exists between NAEP score results and number of television hours watched.[65]

Students were asked how much television (including videotapes) they usually watched each school day. The results for fourth-grade DDESS students are shown in Table 6.3.

- Among fourth graders watching six hours or more, the proportion of DDESS students (20 percent) was not significantly different than at the national level (21 percent).

- The average science scale score for DDESS fourth-grade students who reported watching six hours or more of television on a school day (148) was higher than that for students nationwide (136).

[64] Martin, M. O., I.V.S. Mullis, A.E. Beaton, E.J. Gonzalez, T.A. Smith, and D.L. Kelly. *Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study.* (Boston: TIMSS International Study Center, 1997).

[65] Campbell, J.R., P.L. Donahue, C.M. Reese, and G.W. Phillips. *NAEP 1994 Reading Report Card for the Nation and the States.* (Washington, DC: National Center for Education Statistics, 1996); Beatty, A.S., C.M. Reese, H.R. Persky, and P. Carr. *NAEP 1994 U.S. History Report Card.* (Washington, DC: National Center for Education Statistics, 1996); Campbell, J.R., C.M. Reese, C. O'Sullivan, and J.A. Dossey. *NAEP 1994 Trends in Academic Progress.* (Washington, DC: National Center for Education Statistics, 1996).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 6.3

Grade 4 Students' Reports on Television Viewing Habits

On a school day, about how many hours do you usually watch TV or videotapes outside of school hours?	Percentage and Average Scale Score	
	DDESS	Nation
1 hour or less	28 (1.4) 154 (2.0)	29 (0.8) 148 (1.2)
2 to 3 hours	34 (1.4) 157 (1.7)	34 (0.7) 153 (1.1)
4 to 5 hours	17 (1.0) 155 (1.8)	16 (0.6) 153 (1.5)
6 hours or more	20 (1.2) 148 (1.7)	21 (0.7) 136 (1.5)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Parental Support

When parents are involved in their children's education, both children and parents are likely to benefit. Research on students at risk has shown that parents' participation in their children's education has more effect on the child's performance than parent income or parent education.[66] Parental involvement is naturally part of the home environment, but it is also increasingly sought in the school.

As part of the NAEP assessment, the principals of participating students were asked about parental involvement in their schools. Table 6.4 presents the results for fourth graders in DDESS schools.

- Combining data from two categories shows that, overall, all or nearly all of the fourth-grade students attended schools where principals characterized parental support as somewhat positive or very positive: 100 percent for DDESS, 97 percent for the nation.

- The average scale score for DDESS fourth graders attending schools where parental support was characterized as somewhat positive (158) was higher than that for the students in comparable schools nationwide (147).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 6.4

Schools' Reports on Parental Support at Grade 4

How would you characterize parental support for student achievement within your school?	Percentage and Average Scale Score	
	DDESS	Nation
Somewhat to very negative	0 (****) *** (****)	3 (1.5) 135 (5.3)
Somewhat positive	47 (0.9) 158 (1.5)	57 (4.8) 147 (1.7)
Very positive	53 (0.9) 152 (1.4)	40 (4.6) 150 (2.1)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). *** Sample size is insufficient to permit a reliable estimate. **** Standard error estimates cannot be accurately determined.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

[66] Office of Educational Research and Improvement. *Mapping out the National Assessment of Title I: The Interim Report — 1996.* (Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education, 1996).

Student Mobility

The United States has long been a nation "on the move." Research indicates that moving more than once or twice during the school year lowers student performance. Students who attend the same school throughout their careers are most likely to graduate, whereas the most mobile of the school populations have the highest rates of failure and dropping out.[67] The effects of high mobility are far-reaching; schools with high mobility rates depress performance even for students who do not move.

To examine the relationship between mobility and science performance, the NAEP assessment asked students how many times since starting first grade they had changed schools due to changes in where they lived. Table 6.5 shows results for fourth-grade DDESS students.

- In terms of student mobility, there was no significant difference in the percentages of fourth graders in DDESS schools (19 percent) or nationwide (22 percent) who reported moving only once since starting first grade. For fourth graders moving two times, the percentage of DDESS students (17 percent) was higher than the percentage of comparably mobile students nationwide (8 percent).

- The average scale scores of DDESS students who moved once (157), twice (153), or three or more times (153) since the first grade were higher than those of their national public school counterparts who moved once (148), twice (141), or three or more times (138).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 6.5

Students' Reports on Mobility

Since you started first grade, how many times have you changed schools, not counting when you were promoted to the next grade?	Percentage and Average Scale Score	
	DDESS	Nation
None	22 (1.1) 153 (1.7)	55 (1.2) 152 (1.2)
One	19 (1.0) 157 (2.5)	22 (1.0) 148 (1.5)
Two	17 (0.8) 153 (2.3)	8 (0.5) 141 (2.4)
Three or more	42 (1.4) 153 (1.8)	15 (0.7) 138 (1.4)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

[67] ERIC Clearinghouse on Urban Education. *Highly Mobile Students: Educational Problems and Possible Solutions.* (New York, NY: ERIC Clearinghouse on urban Education, ERIC/CUE Digest, Number 73, 1991).

Students' Views About Science

Science educators have been interested in the relationship between student attitude and student performance for several decades now. A considerable body of research has shown a correlation between students attitudes and their performance in science, with positive attitudes typically resulting in higher performance.[68] Therefore, the 1996 NAEP science assessment asked several questions to gauge students' attitudes towards science. Table 6.6 shows the responses for fourth graders to both a positive and a negative statement about science.

- In DDESS schools, 37 percent of fourth graders agreed that science is useful for solving everyday problems, about the same as at the national level (34 percent). The average scale score for these DDESS students (156) was greater than that for comparable students in the nation (149).

- In DDESS schools, 43 percent of students agreed with the statement that learning science is mostly memorizing facts. The percentage of students in the nation who also held that attitude (40 percent) was not significantly different. The average scale score for DDESS fourth graders (152) who felt that learning science is mostly memorizing was higher than the average scale score of students nationwide (144) who also held that opinion about science.

[68] Weinburg, M. "Gender Differences in Student Attitudes Toward Science: A Meta Analysis of the Literature from 1970 to 1991," in *Journal of Research in Science Teaching, 1985, 32.* pp. 387-398.

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE 6.6

Grade 4 Students' Views About Science

How much do you agree with the following statements?	Percentage and Average Scale Score	
	DDESS	Nation
Science is useful for solving everyday problems.		
Disagree	32 (1.1) 151 (1.4)	32 (0.8) 148 (1.3)
Not sure	32 (1.4) 154 (1.8)	34 (0.7) 149 (1.2)
Agree	37 (1.2) 156 (2.0)	34 (0.8) 149 (1.2)
Learning science is mostly memorizing.		
Disagree	20 (1.1) 157 (1.7)	23 (0.7) 152 (1.1)
Not sure	38 (1.3) 155 (1.4)	37 (0.9) 151 (1.3)
Agree	43 (1.1) 152 (1.7)	40 (0.8) 144 (1.1)

The NAEP science scale ranges from 0 to 300. The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± 2 standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

APPENDIX A

Reporting NAEP 1996 Science Results for DoDEA Schools at Grade 4

The DoDEA schools were assessed at grade 8 as part of the NAEP 1996 science state assessment. The DoDEA arranged to assess its grade 4 students at the same time, although grade 4 was not included in the state science assessment. The grade 4 assessment of DoDEA students was in most ways operationally identical to the national assessment. Appendices A through C, originally written for the state reports, have been rewritten to reflect this.

A.1 Participation Guidelines

As was discussed in the Introduction, unless the overall participation rate for a jurisdiction is sufficiently high, the assessment results for that jurisdiction may be subject to appreciable nonresponse bias. Moreover, even if the overall participation rate is high, significant nonresponse bias may exist if the nonparticipation that does occur is heavily concentrated among certain types of schools or students. The following guidelines concerning school and student participation rates in the state assessment program were established to address four significant ways in which nonresponse bias could be introduced into the jurisdiction sample estimates. For DoDEA schools reported as jurisdictions (as in this report), the guidelines for public schools apply.

The first three guidelines describe the determination of whether a jurisdiction is eligible to have its results published. Guidelines 4-11 describe conditions under which a jurisdiction's published results will include a notation. Such a notation would indicate the possibility of bias in particular results, due to nonresponse from segments of the sample. Note that in order for a jurisdiction's results to be published without notations, that jurisdiction must comply with all guidelines. (A thorough discussion of the NAEP participation guidelines can be found in the *Technical Report of the NAEP 1996 State Assessment Program in Science*.)

Guidelines on the Publication of NAEP Results

Guideline 1 — Publication of Public School Results

A jurisdiction will have its public school results published in the *NAEP 1996 Science Report Card* (or in other reports that include all state-level results) if and only if its weighted participation rate for the initial sample of public schools is greater than or equal to 70 percent. Similarly, a jurisdiction will receive a separate *NAEP 1996 Science State Report* if and only if its weighted participation rate for the initial sample of public schools is greater than or equal to 70 percent.

Guideline 2 — Publication of Nonpublic School Results

A jurisdiction will have its nonpublic school results published in the *NAEP 1996 Science Report Card* (or in other reports that include all state-level results) if and only if its weighted participation rate for the initial sample of nonpublic schools is greater than or equal to 70 percent **AND** meets minimum sample size requirements.[1] A jurisdiction eligible to receive a separate *NAEP 1996 Science State Report* under guideline 1 will have its nonpublic school results included in that report if and only if that jurisdiction's weighted participation rate for the initial sample of nonpublic schools is greater than or equal to 70 percent **AND** meets minimum sample size requirements. If a jurisdiction meets guideline 2 but fails to meet guideline 1, a separate *NAEP 1996 Science State Report* will be produced containing only nonpublic school results.

Guideline 3 — Publication of Combined Public and Nonpublic School Results

A jurisdiction will have its combined results published in the *NAEP 1996 Science Report Card* (or in other reports that include all state-level results) if and only if both guidelines 1 and 2 are satisfied. Similarly, a jurisdiction eligible to receive a separate *NAEP 1996 Science State Report* under guideline 1 will have its combined results included in that report if and only if guideline 2 is also met.

Guidelines for Notations of NAEP Results

Guideline 4 — Notation for Overall Public School Participation Rate

A jurisdiction that meets guideline 1 will receive a notation if its weighted participation rate for the initial sample of public schools was below 85 percent **AND** the weighted public school participation rate after substitution was below 90 percent.

Guideline 5 — Notation for Overall Nonpublic School Participation Rate

A jurisdiction that meets guideline 2 will receive a notation if its weighted participation rate for the initial sample of nonpublic schools was below 85 percent **AND** the weighted nonpublic school participation rate after substitution was below 90 percent.

[1] Minimum participation size requirements for reporting nonpublic school data consist of two components: (1) a school sample size of six of more participating schools and (2) an assessed student sample size of at least 62.

Guideline 6 — Notation for Strata-Specific Public School Participation Rate

A jurisdiction that is not already receiving a notation under guideline 4 will receive a notation if the sample of public schools included a class of schools with similar characteristics that had a weighted participation rate (after substitution) of below 80 percent, and from which the nonparticipating schools together accounted for more than five percent of the jurisdiction' s total weighted sample of public schools. The classes of schools from each of which a jurisdiction needed minimum school participation levels were determined by degree of urbanization, minority enrollment, and median household income of the area in which the school is located.

Guideline 7 — Notation for Strata-Specific Nonpublic School Participation Rate

A jurisdiction that is not already receiving a notation under guideline 5 will receive a notation if the sample of nonpublic schools included a class of schools with similar characteristics that had a weighted participation rate (after substitution) of below 80 percent, and from which the nonparticipating schools together accounted for more than five percent of the jurisdiction's total weighted sample of nonpublic schools. The classes of schools from each of which a jurisdiction needed minimum school participation levels were determined by type of nonpublic school (Catholic versus non-Catholic) and location (metropolitan versus nonmetropolitan).

Guideline 8 — Notation for Overall Student Participation Rate in Public Schools

A jurisdiction that meets guideline 1 will receive a notation if the weighted student response rate within participating public schools was below 85 percent.

Guideline 9 — Notation for Overall Student Participation Rate in Nonpublic Schools

A jurisdiction that meets guideline 2 will receive a notation if the weighted student response rate within participating nonpublic schools was below 85 percent.

Guideline 10—Notation for Strata-Specific Student Participation Rates in Public Schools

A jurisdiction that is not already receiving a notation under guideline 8 will receive a notation if the sampled students within participating public schools included a class of students with similar characteristics that had a weighted student response rate of below 80 percent, and from which the nonresponding students together accounted for more than five percent of the jurisdiction's weighted assessable public school student sample. Student groups from which a jurisdiction needed minimum levels of participation were determined by the age of the student, whether or not the student was classified as a student with a disability (SD) or of limited English proficiency (LEP), and the type of assessment session (monitored or unmonitored), as well as school level of urbanization, minority enrollment, and median household income of the area in which the school is located.

Guideline 11 — Notation for Strata-Specific Student Participation Rates in Nonpublic Schools

A jurisdiction that is not already receiving a notation under guideline 9 will receive a notation if the sampled students within participating nonpublic schools included a class of students with similar characteristics that had a weighted student response rate of below 80 percent, and from which the nonresponding students together accounted for more than five percent of the jurisdiction's weighted assessable nonpublic school student sample. Student groups from which a jurisdiction needed minimum levels of participation were determined by the age of the student, whether or not the student was classified as a student with a disability (SD) or of limited English proficiency (LEP), and the type of assessment session (monitored or unmonitored), as well as type and location of school.

A.2 NAEP Reporting Groups

The NAEP assessment program provides results for groups of students defined by shared characteristics — region of the country, gender, race/ethnicity, parental education, type of school, and participation in federally funded Title I programs and the free/reduced-price lunch component of the National School Lunch Program. (Region of the country and type of school are not applicable to DoDEA schools and hence are not included here, but there are descriptions in the grade 8 DoDEA science state assessment reports.)

Based on criteria described later in this appendix, results are reported for subpopulations only when sufficient numbers of students and adequate school representation are present. For public school students, there must be at least 62 students in a particular subgroup from at least 5 primary sampling units (PSUs).[2] For nonpublic school students, the minimum requirement is 62 students in a particular subgroup from at least 6 different schools. However, the data for all students, regardless of whether their subgroup was reported separately, were included in computing overall results for DoDDS or DDESS. Definitions of the subpopulations referred to in this report are presented on the following pages.

Gender

Results are reported separately for males and females.

Race/Ethnicity

The racial/ethnic results presented in this report attempt to provide a clear picture based on several sources. The race/ethnicity variable is an imputed definition of race/ethnicity derived from up to three sources. This variable is used for race/ethnicity subgroup comparisons. Two questions from the student demographics questionnaire were used in the determination of derived race/ethnicity:

> If you are Hispanic, what is your Hispanic background?
>
> - I am not Hispanic.
> - Mexican, Mexican American, or Chicano
> - Puerto Rican
> - Cuban
> - Other Spanish or Hispanic Background

Students who responded to this question by filling in the second, third, fourth, or fifth oval were considered Hispanic. For students who filled in the first oval, did not respond to the

[2] For the DDESS and DoDDS, a PSU is most often a single school (as it is for the jurisdictions in the state assessments); for the national assessment, a PSU is a selected geographic region (a county, group of counties, or a metropolitan statistical area).

question, or provided information that was illegible or could not be classified, responses to the question below were examined in an effort to determine race/ethnicity.

Which best describes you?

- White (not Hispanic)
- Black (not Hispanic)
- Hispanic ("Hispanic" means someone who is from a Mexican, Mexican American, Chicano, Puerto Rican, Cuban, or other Spanish or Hispanic background.)
- Asian or Pacific Islander ("Asian or Pacific Islander" means someone who is from a Chinese, Japanese, Korean, Filipino, Vietnamese, or other Asian or Pacific Island background.)
- American Indian or Alaskan Native ("American Indian or Alaskan Native" means someone who is from one of the American Indian tribes, or one of the original people of Alaska.)
- Other (specify) ______________________

Students' race/ethnicity was then assigned on the basis of their response. For students who filled in the sixth oval ("Other") or provided illegible information or information that could not be classified, or did not respond at all, race/ethnicity was assigned as determined by school records.[3]

Derived race/ethnicity could not be determined for students who did not respond to either of the demographic questions and for whom a race/ethnicity designation was not provided by the school.

The details of how race/ethnicity classifications are derived is presented so that the readers can determine the usefulness of the results for their particular uses. It should be noted that a nonnegligible number of students indicated a Hispanic background (e.g., Puerto Rican or Cuban) **and** indicated that a racial/ethnic category other than Hispanic best described them. These students were classified as Hispanic according to the rules described above. Also, information from the schools did not always correspond to students' descriptions of themselves.

[3] The procedure for assigning race/ethnicity was modified for Hawaii. See the *Technical Report for the NAEP 1996 State Assessment Program in Science* for details.

Parents' Highest Level of Education

The variable representing level of parental education is derived from responses to two questions from the set of general background questions. Students were asked to indicate the extent of their mothers' education:

> How far in school did your mother go?
>
> - She did not finish high school.
> - She graduated from high school.
> - She had some education after high school.
> - She graduated from college.
> - I don't know.

Students were asked a similar question about their fathers' education:

> How far in school did your father go?
>
> - He did not finish high school.
> - He graduated from high school.
> - He had some education after high school.
> - He graduated from college.
> - I don't know.

This information was combined into one parental education reporting variable through the following procedure. If a student indicated the extent of education for only one parent, that level was included in the data. If a student indicated the extent of education for both parents, the higher of the two levels was included in the data. For students who did not know the level of education for both parents or did not know the level for one parent and did not respond for the other, the parental education level was classified as "I don't know." If the student did not respond for either parent, the student was recorded as having provided no response.

It should be noted that, nationally, approximately one-third of fourth graders reported not knowing the education level of either of their parents.

Title I Participation

On the basis of available school records, students were classified either as currently participating in a Title I program or receiving Title I services, or as not receiving such services.

The classification only refers to the school year when the assessment was administered (i.e., the 1995–96 school year) and is not based on participation in previous years. If the school did not offer any Title I programs or services, all students in that school were classified as not participating.

Free/Reduced-Price School Lunch Program Eligibility

On the basis of available school records, students were classified either as currently eligible or not eligible for the free or reduced-price component of the Department of Agriculture's school lunch program. The classification refers only to the school year when the assessment was administered (i.e., the 1995–96 school year) and is not based on eligibility in previous years. If the school did not participate in the program or if school records were not available, all students in that school were classified as "Information not available."

A.3 Guidelines for Analysis and Reporting

This report describes science performance for fourth graders and compares the results for various groups of students within this population — for example, those who have certain demographic characteristics or who responded to a specific background question in a particular way. The report examines the results for individual demographic groups and individual background questions. It does not include an analysis of the relationships among combinations of these subpopulations or background questions.

Drawing Inferences from the Results

Because the percentages of students in these subpopulations and their average scale scores are based on samples — rather than on the entire population of fourth graders in a jurisdiction — the numbers reported are necessarily *estimates*. As such, they are subject to a measure of uncertainty, reflected in the *standard error* of the estimate. When the percentages or average scale scores of certain groups are compared, it is essential to take the standard error into account, rather than to rely solely on observed similarities or differences. Therefore, the comparisons discussed in this report are based on *statistical tests* that consider both the magnitude of the difference between the averages or percentages and the standard errors of those statistics.

One of the goals of the science assessment program is to estimate scale score distributions and percentages of students in the categories described in A.2 for the overall populations of fourth-grade students in each participating jurisdiction based on the particular samples of students assessed. The use of *confidence intervals*, based on the standard errors, provides a way to make inferences about the population average scale scores and percentages in a manner that reflects the uncertainty associated with the sample estimates. An estimated sample average scale score ± 2 standard errors approximates a *95 percent confidence interval* for the corresponding population average or percentage. This means that one can conclude with approximately 95 percent confidence that the average scale score of the entire population of interest (e.g., all fourth-grade students in public schools in a jurisdiction) is within ± 2 standard errors of the sample average.

As an example, suppose that the average science scale score of the students in a particular jurisdiction's fourth-grade sample were 156 with a standard error of 1.2. A 95 percent confidence interval for the population average would be as follows:

$$\text{Average} \pm 2 \text{ standard errors} = 156 \pm 2 \times (1.2) = 156 \pm 2.4 =$$
$$156 - 2.4 \text{ and } 156 + 2.4 = (153.6, 158.4)$$

Thus, one can conclude with 95 percent confidence that the average scale score for the entire population of fourth-grade students in public schools in that jurisdiction is between 153.6 and 158.4.

Similar confidence intervals can be constructed for percentages, *if the percentages are neither extremely large nor extremely small.* For extreme percentages, confidence intervals constructed in the above manner may not be appropriate, and accurate confidence intervals can be constructed only by using procedures that are quite complicated.

Extreme percentages, defined by both the magnitude of the percentage and the size of the sample from which it was derived, should be interpreted with caution. (The forthcoming *Technical Report of the NAEP 1996 State Assessment Program in Science* contains a more complete discussion of extreme percentages.)

Analyzing Subgroup Differences in Averages and Percentages

The statistical tests determine whether the evidence, based on the data from the groups in the *sample*, is strong enough to conclude that the averages or percentages are actually different for those groups in the *population.* If the evidence is strong (i.e., the difference is statistically significant), the report describes the group averages or percentages as being different (e.g., one group performed *higher than* or *lower than* another group), regardless of whether the sample averages or sample percentages appear to be about the same or not. If the evidence is not sufficiently strong (i.e., the difference is not statistically significant), the averages or percentages are described as being *not significantly different* — again, regardless of whether the sample averages or sample percentages appear to be about the same or widely discrepant. When determining whether sample differences are likely to represent actual differences between the groups in the population, the results of the statistical tests should be relied on rather than the apparent magnitude of the difference between sample averages or percentages.

In addition to the overall results, this report presents outcomes separately for a variety of important subgroups. Many of these subgroups are defined by shared characteristics of students, such as their gender or race/ethnicity. Other subgroups are defined by the responses of the assessed students' science teachers to questions in the science teacher questionnaire.

In Chapter 1 of this report, differences between the jurisdiction and the nation were tested for overall science scale score and for each of the fields of science. In Chapter 2, significance tests were conducted for the overall scale score for each of the subpopulations. In Chapters 3 through 6, comparisons were made across subgroups for responses to various background questions.

As an example of comparisons across subgroups, consider the question: *Do students who reported discussing studies at home almost every day exhibit higher average science scale scores than students who report never or hardly ever doing so?*

To answer the above question, begin by comparing the average science scale score for the two groups being analyzed. If the average for the group that reported discussing their studies at home almost every day is higher, it may be tempting to conclude that that group does have a higher science scale score than the group that reported never or hardly ever discussing their studies at home. However, even though the averages differ, there may be no real difference in performance between the two groups in the population because of the uncertainty associated with the estimated average scale scores of the groups in the sample. Remember that the intent is to make a statement about the entire population, not about the particular sample that was assessed. The data from the sample are used to make inferences about the population as a whole.

As discussed in the previous section, each estimated sample average scale score (or percentage) has a degree of uncertainty associated with it. It is therefore possible that if all students in the population (rather than a sample of students) had been assessed or if the assessment had been repeated with a different sample of students or a different, but equivalent, set of questions, the performances of various groups would have been different. Thus, to determine whether there is a *real* difference between the average scale score (or percentage of students with a certain attribute) for two groups in the population, an estimate of the degree of uncertainty associated with the difference between the scale score averages or percentages of those groups must be obtained for the sample. This estimate of the degree of uncertainty — called *the standard error of the difference* between the groups — is obtained by taking the square of each group's standard error, summing these squared standard errors, and then taking the square root of this sum.

In a manner similar to that in which the standard error for an individual group average or percentage is used, the *standard error of the difference* can be used to help determine whether differences between groups in the population are real. The difference between the mean scale score or percentage of the two groups ± 2 standard errors of the difference represents an approximate 95 percent confidence interval. If the resulting interval includes zero, there is insufficient evidence to claim a real difference between groups in the population. If the interval does not contain zero, the difference between groups is *statistically significant* (different) at the 0.05 level.

As another example, to determine whether the average science scale score of fourth-grade males is higher than that of fourth-grade females in a particular jurisdiction's public schools, suppose that the sample estimates of the average scale scores and standard errors for males and females were as follows:

Group	Average Scale Score	Standard Error
Males	148	0.9
Females	146	1.1

The difference between the estimates of the average scale scores of males and females is two points (148 −146). The standard error of this difference is

$$\sqrt{0.9^2 + 1.1^2} = 1.4$$

Thus, an approximate 95 percent confidence interval for this difference is

Mean difference ± 2 standard errors of the difference =
$2 \pm 2 \times (1.4) = 2 \pm 2.8 = 2 - 2.8 \text{ and } 2 + 2.8 = (-0.8, 4.8)$

The value zero is within this confidence interval, which extends from − 0.8 to 4.8 (i.e., zero is between − 0.8 and 4.8). Thus, there is insufficient evidence to claim a difference in average science scale score between the populations of fourth-grade males and females in public schools in the hypothetical jurisdiction.

Throughout this report, when the average scale scores or percentages for two groups were compared, procedures like the one described above were used to draw the conclusions that are presented in the text.[4] If a statement appears in the report indicating that a particular group had a *higher* (or *lower*) average scale score than a second group, the 95 percent confidence interval for the difference between groups did not contain zero. An attempt was made to distinguish between group differences that were statistically significant but rather small in a practical sense and differences that were both statistically and practically significant. A procedure based on effect sizes was used. Statistically significant differences that are rather small are described in the text as *somewhat higher* or *somewhat lower*. When a statement indicates that the average scale score or percentage of some attribute was *not significantly different* for two groups, the confidence interval included zero, and thus no difference could be inferred between the groups. The reader is cautioned to avoid drawing conclusions solely on the basis of the magnitude of the difference. A difference between two groups in the sample that appears to be slight may represent a statistically significant difference in the population because of the magnitude of the standard errors. Conversely, a difference that appears to be large may not be statistically significant.

[4] The procedure described above (especially the estimation of the standard error of the difference), is, in a strict sense, only appropriate when the statistics being compared come from independent samples. For certain comparisons in the report, the groups were not independent. In those cases, a different (and more appropriate) estimate of the standard error of the difference was used.

The procedures described in this section, and the certainty ascribed to intervals (i.e., a 95 percent confidence interval), are based on statistical theory that assumes that only one confidence interval or test of statistical significance is being performed. However, in each chapter of this report, many different groups are being compared (i.e., multiple sets of confidence intervals are being calculated). In sets of confidence intervals, statistical theory indicates that the certainty associated with the entire set of intervals is less than that attributable to each individual comparison from the set if considered individually. To hold the certainty level for the set of comparisons at a particular level (i.e., 0.95), modifications (called multiple comparison procedures) must be made to the methods described in the previous section. One such procedure — the *Bonferroni* method — was used in the analyses described in this report to form confidence intervals for the differences between groups whenever sets of comparisons were considered.[5] Using this method, the confidence intervals in the text that are based on sets of comparisons are more conservative than those described on the previous pages. In other words, some comparisons that were individually statistically significant using the methods previously described may not be statistically significant when the Bonferroni method was used to take the number of related comparisons into account.

Most of the multiple comparisons in this report pertain to relatively small sets or "families" of comparisons. For example, when comparisons were discussed concerning students' reports of parental education, six comparisons were conducted — all pairs of the four parental education levels. In these situations, Bonferroni procedures were appropriate. However, consider another example in Chapter 1 of the grade 8 DoDEA reports: these reports contain a map comparing DoDDS or DDESS average scores with those of the 43 other jurisdictions reporting public school results for the state assessment. To control the certainty level for a large family of comparisons such as this (43), the false discovery rate (FDR) criterion[6] was used. Unlike the Bonferroni procedures which control the familywise error rate (i.e., the probability of making even one false rejection in the set of comparisons), the Benjamini and Hochberg (BH) approach using the FDR criterion controls the expected proportion of falsely rejected hypotheses as a proportion of all rejected hypotheses. Bonferroni procedures may be considered conservative for large families of comparisons.[7] In other words, using the Bonferroni method would produce more statistically nonsignificant comparisons than using the BH approach. A more detailed description of the Bonferroni and BH procedures appears in the *Technical Report of the NAEP 1996 State Assessment Program in Science.*

[5] Miller, R.G. *Simultaneous Statistical Inference.* (New York, NY: McGraw-Hill, 1966).

[6] Benjamini, Y. and Hochberg. "Controlling the false discovery rate: A practical and powerful approach to multiple testing," in *Journal of the Royal Statistical Society, Series B,* 57(1). (pp. 289-300, 1994).

[7] Williams, V.S.L., L.V. Jones, and J.W. Tukey. *Controlling Error in Multiple Comparisons, with Special Attention to the National Assessment of Educational Progress.* (Research Triangle Park, NC: National Institute of Statistical Sciences, December 1994).

Statistics with Poorly Estimated Standard Errors

Not only are the averages and percentages reported in NAEP subject to uncertainty, but their standard errors are as well. In certain cases, typically when the standard error is based on a small number of students or when the group of students is enrolled in a small number of schools, the amount of uncertainty associated with the standard errors may be quite large. Throughout this report, estimates of standard errors subject to a large degree of uncertainty are followed by the symbol "!". In such cases, the standard errors — and any confidence intervals or significance tests involving these standard errors — should be interpreted cautiously. Additional details concerning procedures for identifying such standard errors are discussed in the *Technical Report of the NAEP 1996 State Assessment Program in Science.*

Minimum Subgroup Sample Sizes

Results for science performance and background variables were tabulated and reported for groups defined by gender, race/ethnicity, parental education, type of school, and participation in federally funded Title I programs and the free or reduced-price school lunch component of the National School Lunch Program. NAEP collects data for five racial ethnic subgroups (White, Black, Hispanic, Asian/Pacific Islander, and American Indian/Alaskan Native) and four levels of parents' education (Graduated From College, Some Education After High School, Graduated From High School, and Did Not Finish High School) plus the category "I Don't Know."

In many jurisdictions, and for some regions of the country, the number of students in some of these groups was not sufficiently high to permit accurate estimation of performance and/or background variable results. As a result, data are not provided for the subgroups with students from very few schools or for the subgroups with very small sample sizes. For results to be reported for any state assessment subgroup, public school results must represent at least 5 primary sampling units (PSUs) and nonpublic school results must represent at least 6 schools. For results to be reported for any national assessment subgroup, at least 5 PSUs must be represented in the subgroup. In addition, a minimum sample of 62 students per subgroup is required. For statistical tests pertaining to subgroups, the sample size for both groups has to meet the minimum sample size requirements.

The minimum sample size of 62 was determined by computing the sample size required to detect an effect size of 0.5 total-group standard deviation units with a probability of 0.8 or greater. The effect size of 0.5 pertains to the *true* difference between the average scale score of the subgroup in question and the average scale score for the total fourth-grade public school population in the jurisdiction, divided by the standard deviation of the scale score in the total population. If the *true* difference between subgroup and total group mean is 0.5 total-group standard deviation units, then a sample size of at least 62 is required to detect such a difference with a probability of 0.8. Further details about the procedure for determining minimum sample size appear in the *Technical Report of the NAEP 1996 State Assessment Program in Science.*

Describing the Size of Percentages

Some of the percentages reported in the text of the report are given qualitative descriptions. For example, the number of students currently taking a biology class might be described as "relatively few" or "almost all," depending on the size of the percentage in question. Any convention for choosing descriptive terms for the magnitude of percentages is to some degree arbitrary. The descriptive phrases used in the report and the rules used to select them are shown below.

Percentage	Descriptive Term Used in Report
$p = 0$	None
$0 < p \leq 8$	A small percentage
$8 < p \leq 13$	Relatively few
$13 < p \leq 18$	Less than one fifth
$18 < p \leq 22$	About one fifth
$22 < p \leq 27$	About one quarter
$27 < p \leq 30$	Less than one third
$30 < p \leq 36$	About one third
$36 < p \leq 47$	Less than half
$47 < p \leq 53$	About half
$53 < p \leq 64$	More than half
$64 < p \leq 71$	About two thirds
$71 < p \leq 79$	About three quarters
$79 < p \leq 89$	A large majority
$89 < p \leq 100$	Almost all
$p = 100$	All

APPENDIX B

The NAEP 1996 Science Assessment

The science framework for the 1996 National Assessment of Educational Progress was produced under the auspices of the National Assessment Governing Board through a consensus process. The consensus process, managed by the Council of Chief State School Officers, with the National Center for Improving Science Education and the American Institutes for Research, developed the framework over a ten-month period between October 1990 and August 1991. The following factors guided the process for developing consensus on the science framework:[8]

- the active participation of individuals such as curriculum specialists, science teachers, science supervisors, state supervisors, administrators, individuals from business and industry, government officials, and parents;
- the representation of what is considered essential learning in science, and the recommendation of innovative assessment techniques to probe the critical abilities and content areas;
- the recognition of the lack of agreement on such things as common scope of instruction and sequence, components of scientific literacy, important outcomes of learning, and the nature of overarching themes in science.

[8] *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1993).

While maintaining some conceptual continuity with the 1990 NAEP Science Assessment, the 1996 framework takes into account the current reforms in science education, as well as documents such as the science framework used for the 1991 International Assessment of Educational Progress. In addition, the Framework Steering Committee recommended that a variety of strategies, including the following, be used for assessing students' performance:[9]

- performance tasks that allow students to manipulate physical objects and draw scientific understanding from the materials before them;
- constructed-response questions that provide insights into students' levels of understanding and ability to communicate in the sciences as well as their ability to generate, rather than simply recognize, information related to scientific concepts and their interconnections; and
- multiple-choice items that probe students' conceptual understanding and ability to connect ideas in a scientifically sound way.

B.1 Percentage of Assessment Time by Domain

The framework for the 1996 science assessment can be described as a two-dimensional matrix. The three fields of science (earth, physical, and life) make up the first dimension and ways of knowing and doing science (conceptual understanding, scientific investigation, and practical reasoning) make up the second dimension. Every question or task in the assessment is classified according to the two major dimensions. There are also two overarching domains—nature of science (that includes nature of technology) and themes (systems, models, and patterns of change).

In addition to describing the content of the assessment, the framework also recommends what percentage of time should be devoted to each field of science, each way of knowing and doing science, the nature of science, and themes.

In this section, each figure describes an element of the framework, and is followed by a table showing the *actual* distribution of assessment time as well as the distribution *recommended* by the framework. Care was taken to ensure congruence between the proportions actually used in the assessment and those recommended in the assessment specifications. Note that the tables represent all three grades assessed nationally; only grade 8 was assessed at the state level.

Figure B.1 describes the fields of science and Table B.1 shows the actual and recommended distribution of assessment time across each field. The ways of knowing and doing science are outlined in Figure B.2. The distribution of assessment time for this dimension, both actual and recommended, is depicted in Table B.2.

---

[9] *Ibid.*

THE NATION'S REPORT CARD naep 1996 State Assessment

FIGURE B.1

Description of the Three Fields of Science

Earth Science

The earth science content assessed centers on objects and events that are relatively accessible or visible. The concepts and topics covered are solid Earth (lithosphere). water (hydrosphere), air (atmosphere), and the Earth in space. The solid Earth consists of composition; forces that alter its surface; the formation, characteristics and uses of rocks; the changes and uses of soil; natural resources used by humankind; and natural forces within the Earth. Concepts and topics related to water consist of the water cycle; the nature of oceans and their effects on water and climate; and the location of water, its distribution, characteristics, and effect of and influence on human activity. The air is broken down into composition and structure of the atmosphere (including energy transfer); the nature of weather; common weather hazards; and air quality and climate. The Earth in space consists of the setting of the Earth in the solar system; the setting and evolution of the solar system in the universe; tools and technology that are used to gather information about space; apparent daily motions of the Sun, the Moon, the planets and the stars; rotation of the Earth about its axis, and the Earth's revolution around the Sun; and tilt of the Earth's axis that produces seasonal variations in the climate.

Physical Science

The physical science component relates to basic knowledge and understanding concerning the structure of the universe as well as the physical principles that operate within it. The major subtopics probed are matter and its transformations, energy and its transformations, and the motion of things. Matter and its transformations are described by diversity of materials (classification and types and the particulate nature of matter); temperature and states of matter; properties and uses of material (modifying properties, synthesis of materials with new properties); and resource management. Energy and its transformations involve different forms of energy; energy transformations in living systems, natural physical systems, and artificial systems constructed by humans; and energy sources and use, including distribution, energy conversion, and energy costs and depletion. Motion is broken down into an understanding of frames of reference; force and changes in position and motion; action and reaction; vibrations and waves as motion; general wave behavior; electromagnetic radiation; and the interactions of electromagnetic radiation with matter.

Life Science

The fundamental goal of life science is to attempt to understand and explain the nature and function of living things. The major concepts assessed in life science are change and evolution, cells and their functions (not at grade 4), organisms, and ecology. Change and evolution includes diversity of life on Earth; genetic variation within a species; theories of adaptation and natural selection; and changes in diversity over time. Cells and their functions consists of information transfer; energy transfer for the construction of proteins; and communication among cells. Organisms are described by reproduction, growth and development; life cycles; and functions and interactions of systems within organisms. The topic of ecology centers on the interdependence of life — populations, communities, and ecosystems.

SOURCE: *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board; 1993).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE B.1

Distribution of Assessment Time by Field of Science

	Earth		Physical		Life	
	Actual	Recommended	Actual	Recommended	Actual	Recommended
Grade 4	33%	33%	34%	33%	33%	33%
Grade 8	30%	30%	30%	30%	40%	40%
Grade 12	33%	33%	33%	33%	34%	33%

THE NATION'S REPORT CARD naep 1996 State Assessment

FIGURE B.2

Description of Knowing and Doing Science

Conceptual Understanding

Conceptual understanding includes the body of scientific knowledge that students draw upon when conducting a scientific investigation or engaging in practical reasoning. Essential scientific concepts involve a variety of information including facts and events the student learns from science instruction and experiences with the natural environment and scientific concepts, principles, laws, and theories that scientists use to explain and predict observations of the natural world.

Scientific Investigation

Scientific investigation probes students' abilities to use the tools of science, including both cognitive and laboratory tools. Students should be able to acquire new information, plan appropriate investigations, use a variety of scientific tools, and communicate the results of their investigations.

Practical Reasoning

Practical reasoning probes students' ability to use and apply science understanding in new, real-world applications.

SOURCE: *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1993).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE B.2

Distribution of Assessment Time by Knowing and Doing Science

	Conceptual Understanding		Scientific Investigation		Practical Reasoning	
	Actual	Recommended	Actual	Recommended	Actual	Recommended
Grade 4	45%	45%	38%	45%	17%	10%
Grade 8	45%	45%	29%	30%	26%	25%
Grade 12	44%	45%	28%	30%	28%	25%

The two overarching dimensions are described and accounted for by Figure B.3 and Table B.3, which describe the nature of science and the themes that transcend the scientific disciplines.

THE NATION'S REPORT CARD naep 1996 State Assessment

FIGURE B.3

Description of Overarching Domains

The Nature of Science

The nature of science incorporates the historical development of science and technology, the habits of mind that characterize these fields, and methods of inquiry and problem-solving. It also encompasses the nature of technology that includes issues of design, application of science to real-world problems, and trade-offs or compromises that need to be made.

Themes

Themes are the "big ideas" of science that transcend the various scientific disciplines and enable students to consider problems with global implications. The NAEP science assessment focuses on three themes: systems, models, and patterns of change.

- Systems are complete, predictable cycles, structures or processes occurring in natural phenomena. Students should understand that a system is an artificial construction created to represent, or explain a natural occurrence. Students should be able to identify and define the system boundaries, identify the components and their interrelationships and note the inputs and outputs to the system.

- Models of objects and events in nature are ways to understand complex or abstract phenomena. As such they have limits and involve simplifying assumptions but also possess generalizability and often predictive power. Students need to be able to distinguish the idealized model from the phenomenon itself and to understand the limitations and simplified assumptions that underlie scientific models.

- Patterns of change involve students' recognition of patterns of similarity and differences, and recognize how these patterns change over time. In addition, students should have a store of common types of patterns and transfer their understanding of a familiar pattern of change to a new and unfamiliar one.

SOURCE: *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1993).

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE B.3

Distribution of Assessment Time by Overarching Domains

	Nature of Science		Themes	
	Actual	Recommended	Actual*	Recommended
Grade 4	19%	≥15%	53%	33%
Grade 8	21%	≥15%	49%	50%
Grade 12	31%	≥15%	55%	50%

* Several of the hands-on tasks were classified as themes.

SOURCE: *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1993).

B.2 The Assessment Design

The DoDEA grade 4 science assessment used booklets that were identical to those used at grade 4 for the national assessment. Each student in the science assessment received a booklet containing six sections. Three of these sections were blocks[10] of cognitive questions that assessed the knowledge and skills outlined in the framework, and the other three sections were sets of background questions. Two of the three cognitive sections were paper-and-pencil, and the third section consisted of a hands-on task with related questions. Students at grade 4 were given cognitive blocks that each required 20 minutes to complete.

There were 15 different sections or blocks of cognitive questions, but each student's booklet contained only three of these blocks of items. Every block consisted of both multiple-choice and constructed-response questions. Short constructed-response questions required a few words or a sentence or two for an answer (e.g., briefly stating how nutrients move from the digestive system to the tissues) while the extended constructed-response questions generally required a paragraph or more (e.g., outlining an experiment to test the effect of increasing the amount of available food on the rate of increase of the hydra population). Some constructed-response questions also required diagrams, graphs, or calculations. It was expected that students could adequately answer the short constructed-response questions in about 2 to 3 minutes and the extended constructed-response questions in about 5 minutes.

Other features were built into the blocks of cognitive questions. Four of the blocks were hands-on tasks in which students were given a set of equipment and asked to conduct an investigation and answer questions relating to it. Every student was assessed on one of these four blocks. A second feature was the inclusion of three theme blocks — one assessing systems, one assessing models, and one assessing patterns of change. For example, students were shown a simplified model of part of the Solar System with a brief description, and then asked a number of questions based on this scenario. Theme blocks were randomly placed in booklets, but not in all booklets. No student received more than one theme block.

Each booklet in the assessment also included three sets of student background questions. The first, consisting of general background questions, asked students about such things as mother's and father's level of education, reading materials in the home, homework, and school attendance. The second, consisting of science background questions, asked students questions about their classroom learning activities such as hands-on exercises, courses taken, use of specialized resources such as computers, and views on the utility and value of science. To complete these two questionnaires, students at all grades were given 5 minutes (with the exception of the general background questionnaire for grade 4 students where more time was necessary because the questions were read aloud to the students). The third background questionnaire contained five questions about students' motivation to do well on the assessment, their perception of the difficulty of the assessment, and their familiarity with the types of cognitive questions asked. This section took 3 minutes or less to complete.

Using information gathered from the field test, the booklets were carefully constructed to balance time requirements for the question types in each block. For more information on the design of the assessment, refer to Appendix C.

[10] "Blocks" are separately-timed collections of questions grouped, in part, according to the amount of time required to answer them.

B.3 Usage of Question Types

The data in Table B.4 reflect the number of questions by type and by grade level for the 1996 assessment. One hundred and sixty-five multiple-choice (MC), 219 short constructed-response (SCR), and 59 extended constructed-response (ECR) questions make up the assessment, giving a total of 443 unique questions in the pool. Some of these questions were used at more than one grade level; thus, the sum at each grade level is greater than the total number of unique questions. For the assessment at grade 4, students responded to subsets (determined by booklet) of 51 multiple-choice questions, 73 short constructed-response questions, and 16 extended constructed-response tasks.

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE B.4

Distribution of Items by Question Type

	Grade 4			Grade 8			Grade 12		
	MC	SRC	ERC	MC	SRC	ERC	MC	SRC	ERC
Grade 4 only	42	57	12						
Grades 4 & 8 overlap	9	16	4	9	16	4			
Grade 8 only				44	58	13			
Grades 8 & 12 overlap				21	26	3	21	26	3
Grade 12 only							49	62	27
TOTAL by grade	51	73	16	74	100	20	70	88	30

MC — multiple-choice questions; SRC — short constructed-response questions; ERC — extended constructed-response questions

APPENDIX C

Technical Appendix: The Design, Implementation, and Analysis of the 1996 Assessment in Science for Grade 4 DoDEA Students

C.1 Overview

In 1996, NAEP included a national science assessment at grades 4, 8, and 12, and a state science assessment at grade 8 only. DoDDS and DDESS were the only separate jurisdictions in which a fourth grade science assessment was conducted. The purpose of this appendix is to provide technical information about the 1996 DoDEA fourth grade assessment in science. It describes the design of the assessment and gives an overview of the steps used to implement the program, from the planning stages through the analysis of the data.

This appendix is one of several documents that provide technical information about the 1996 assessment program. Additional details are in the *NAEP 1996 Technical Report* and the *Technical Report of the NAEP 1996 State Assessment Program in Science*. Theoretical information about the models and procedures used in NAEP can be found in the special NAEP-related issue of the *Journal of Educational Statistics* (Summer 1992/Volume 17, Number 2) as well as previous national technical reports.

Educational Testing Service (ETS) was awarded the cooperative agreement for the 1996 NAEP programs, including the DoDEA assessments. ETS was responsible for overall management of the programs as well as for development of the overall design, the cognitive questions and questionnaires, data analysis, and reporting. National Computer Systems (NCS) was a subcontractor to ETS on both the national and state NAEP programs. NCS was responsible for printing, distributing, and receiving all assessment materials, and for scanning and scoring the assessments. The National Center for Education Statistics (NCES) awarded a separate cooperative agreement to Westat, Inc., for handling all aspects of sampling and field operations for the national, state, and fourth-grade DoDEA assessments for 1996.

Organization of the Technical Appendix

This appendix has the following organization:

- Section C.2 provides an overview of the design of the 1996 assessment in science for DoDEA schools.
- Section C.3 discusses the partially-balanced incomplete block (PBIB) spiral design used to assign cognitive questions to assessment booklets and assessment booklets to students.
- Section C.4 outlines the sampling design used for the 1996 assessment.
- Section C.5 summarizes Westat's field administration procedures.
- Section C.6 describes the flow of the data from receipt at NCS through data entry and professional scoring.
- Section C.7 summarizes the procedures used to weight the assessment data and to obtain estimates of the sampling variability of subpopulation estimates.
- Section C.8 describes the initial analyses performed to verify the quality of the data.
- Section C.9 describes the item response theory scales and the overall science composite scale created for the final analyses of the data.
- Section C.10 provides an overview of the linking of the DoDEA grade 4 science results to those from the national assessment.

C.2 Design of the NAEP 1996 Assessment in Science for DoDEA Schools

The design for the assessments in science included the following major aspects:

- The fourth-grade science assessment instruments used for the DoDEA assessments program and the national assessment consisted of 15 blocks of questions, of which 4 were hands-on tasks. Each block could contain a mixture of question types — constructed-response or multiple-choice — that was determined by the nature of the task. In addition, the constructed-response questions were of two types: *short constructed-response* questions required students to respond to a question with a few words or a few sentences, while *extended constructed-response* questions required students to respond to a question with a paragraph or more, sometimes including graphs or calculations. The hands-on tasks were similar to laboratory exercises. Each student was given 2 of the 11 cognitive blocks of questions, and one of the 4 hands-on blocks.
- A complex form of matrix sampling called a partially balanced incomplete block (PBIB) spiraling design was used. With PBIB spiraling, students in an assessment session received different booklets containing 3 of the 15 blocks. This provided for

greater science content coverage without the undue testing burden that would have resulted from administering the full set of questions to each student.

- Sets of background questions given to the students, the students' science teachers, and the principals or other school administrators provided a variety of contextual information. The background questionnaires for the DoDEA assessments were identical to those used in the national fourth-grade assessment.

- The total assessment time for each student was approximately two hours, including cleanup and collection of materials from hands-on tasks. Each assessed fourth-grade student was assigned a science booklet that contained 3 of the 15 blocks of science questions requiring 20 minutes each (including a hands-on task block in the last position), followed by a 5-minute general background questionnaire (with additional time for the administrator to read each question), a 5-minute science background questionnaire, and a 3-minute motivation questionnaire. Thirty-seven different booklets were assembled.

- The assessments were administered in the five-week period between January 29 and March 4, 1996. One-fourth of the schools in each jurisdiction were assessed each week throughout the first four weeks. Because of the severe weather throughout much of the country, the fifth week was used for regular testing as well as for makeup sessions.

To assure that the assessment was administered under standard, uniform procedures, data collection at DoDEA schools employed the same methods that were used for the national sample. Security and uniform assessment administration were high priorities. For both DDESS and DoDDS, the presence of Westat staff members, who were on site administering the national assessment at the same time, provided that the grade 4 science assessment was held to the same standards as the national assessment.

C.3 Assessment Instruments

The *student assessment booklets* contained six sections and included both cognitive and noncognitive questions. The assembly of cognitive questions into booklets and their subsequent assignment to assessed students were determined by a matrix sampling design using a variant of a balanced incomplete block design (BIB), with spiraled administration. Each assessed student received a booklet containing 3 of the 15 cognitive blocks according to a design that ensured that each block was administered to a representative sample of students within each jurisdiction. The third cognitive block was always one of the four hands-on blocks; this requirement meant that the BIB was partially balanced (PBIB).

For grade 4, in addition to two 20-minute sections of cognitive questions and the 20-minute performance task section, each booklet included two 5-minute sets of general[11] and science background questions designed to gather contextual information about students, their experiences in science, and their attitudes toward the subject, and one 3-minute section of

[11] The general background questions took longer than 5 minutes for fourth graders, because each question was read aloud by the administrator.

motivation questions designed to gather information about the student's level of motivation while taking the assessment.

In addition to the student assessment booklets, three other instruments provided data relating to the assessment: a science teacher questionnaire, a school characteristics and policies questionnaire, and an SD/LEP student questionnaire (for students categorized as students with disabilities or with limited English proficiency).

The *teacher questionnaire* was administered to the science teachers of the fourth-grade students participating in the assessment. The questionnaire consisted of three sections and took approximately 20 minutes to complete. The first section focused on the teacher's general background and experience; the second, on the teacher's background related to science; and the third, on classroom information about science instruction.

The *school characteristics and policies questionnaire* was given to the principal or other administrator in each participating school and took about 20 minutes to complete. The questions asked about the principal's background and experience, school policies, programs, and facilities, and the demographic composition and background of the students and teachers.

The SD/LEP *student questionnaire* was completed by the staff member most familiar with any student selected for the assessment who was classified in either of two ways: students with disabilities (SD) who had an Individualized Education Plan (IEP) or equivalent special education plan (for reasons other than being gifted and talented); students with limited English proficiency were classified as LEP students. The questionnaire took approximately 3 minutes to complete and asked about the student and the special programs in which the student participated. It was completed for all selected SD or LEP students regardless of whether or not they participated in the assessment. Selected SD or LEP students participated in the assessment if they were determined by the school to be able to participate, considering the terms of their IEP and accommodations provided by the school or by NAEP.

C.4 The Sampling Design

The sampling design for NAEP is complex, in order to minimize burden on schools and students while maximizing the utility of the data. For additional details see the *NAEP 1996 Technical Report*. The target populations for the science assessment reported here consisted of fourth-grade students enrolled in either domestic or overseas DoDEA schools. The representative samples of fourth graders came from 39 DDESS schools or 91 DoDDS schools.

The school samples in DDESS or DoDDS were designed to produce aggregate estimates for the jurisdiction and for selected subpopulations (depending upon the size and distribution of the various subpopulations within the jurisdiction) and to ensure comparability with the national sample.

The national results cited in this report are based on nationally representative samples of fourth-grade students. The samples were selected using a complex multistage sampling design involving the sampling of students from selected schools within selected geographic areas across the country. The sample design had the following stages:

(1) selection of geographic areas (a county, group of counties, or a metropolitan statistical area);

(2) selection of schools (public and nonpublic) within the selected areas; and

(3) selection of students within selected schools.

Each selected school that participated in the assessment and each student assessed represent a portion of the population of interest. To make valid inferences from student samples to the respective populations from which they were drawn, sampling weights are needed. Discussions of sampling weights and how they are used in analyses are presented in sections C.7 and C.8.

Because the fourth-grade DoDEA science samples were too small for precise estimation of item parameters, no scaling was conducted on the sample data. Rather, the parameters for the national fourth-grade sample were used in analyses of the DoDEA data. This facilitates the comparison between the DoDEA results and national results because it places them on the same scale without requiring any additional transformations.

C.5 Field Administration

Administering the 1996 program required collaboration among staff in the participating jurisdictions and schools and the NAEP contractors, especially Westat, the field administration contractor. Details are available in the *NAEP 1996 Technical Report.*

C.6 Materials Processing, Professional Scoring, and Database Creation

Upon completion of each assessment session, school personnel shipped the assessment booklets and forms to NCS for professional scoring, entry into computer files, and checking. The files were then sent to ETS for creation of the database.

After NCS received all appropriate materials from a school, they were forwarded to the professional scoring area where the responses to constructed-response questions were evaluated by trained staff members using guidelines prepared by ETS. Each constructed-response question had a unique scoring guide that defined the criteria to be used in evaluating students' responses. The extended constructed-response questions were evaluated with four- or five-level rubrics. Some of the short constructed-response questions were rated according to three-level rubrics that permit partial credit to be given; other short constructed-response questions were scored as either acceptable or unacceptable.

For the national science assessment and the state assessment program in science, over 4.1 million constructed responses were scored. This figure includes rescoring to monitor interrater reliability. The overall percentage of agreement between scorers for the reliability sample was 93 percent for the tasks in the cognitive blocks and 95 percent for the hands-on tasks.

Data transcription and editing procedures were used to generate the disk and tape files containing various assessment information, including the sampling weights required to make valid statistical inferences about the population from which the DoDEA sample was drawn. Prior to analysis, the data from these files underwent a quality control check at ETS. The files were then merged into a comprehensive, integrated database.

C.7 Weighting and Variance Estimation

A complex sample design was used to select the students who were assessed. The properties of a sample selected through a complex design are very different from those of a simple random sample in which every student in the target population has an equal chance of selection and in which the observations from different sampled students can be considered to be statistically independent of one another. Therefore, the properties of the sample for the complex state assessment program design were taken into account during the analysis of the assessment data.

One way that the properties of the sample design were addressed was by using sampling weights to account for the fact that the probabilities of selection were not identical for all students. All population and subpopulation characteristics based on the assessment data used sampling weights in their estimation. These weights included adjustments for school and student nonresponse.

Not only must appropriate estimates of population characteristics be derived, but appropriate measures of the degree of uncertainty must be obtained for those statistics. One component of uncertainty results from sampling variability, which is a measure of the dependence of the results on the particular sample of students actually assessed. Because of the effects of cluster selection (schools are selected first, then students are selected within those schools), observations made on different students cannot be assumed to be independent of each other (and, in fact, are generally positively correlated). As a result, classical variance estimation formulas will produce incorrect results. Thus, a jackknife variance estimation procedure that accounts for the characteristics of the sample was used for all analyses.

Jackknife variance estimation provides a reasonable measure of uncertainty for any statistic based on values observed without error. Statistics such as the percentage of students correctly answering a given question meet this requirement, but other statistics based on estimates of student science performance, such as the average science scale score of a subpopulation, do not. Because each student typically responds to relatively few questions from a particular field of science (e.g., physical or life science), a nontrivial amount of imprecision exists in the measurement of the scale score of a given student. This imprecision adds another component of variability to statistics based on estimates of individual performance.

C.8 Preliminary Data Analysis

After the computer files of student responses were received and merged into an integrated database, all cognitive and noncognitive questions were subjected to an extensive item analysis. For each cognitive question, this analysis yielded the number of respondents, the percentage of responses in each category, the percentage who omitted the question, the percentage who did not reach the question, and the correlation between the question score and the block score. In addition, the item analysis program provided summary statistics for each block of cognitive questions, including a reliability (internal consistency) coefficient. These analyses were used to check the scoring of the questions, to verify that the difficulty level of the questions was appropriate, and to ensure that students had received adequate time to complete the assessment. The results were reviewed by knowledgeable project staff members in search of aberrations that might signal unusual results or errors in the database.

C.9 Scaling the Assessment Questions

The primary analysis and reporting of the results from the national assessment program used item response theory (IRT) scale-score models. Scaling models quantify a respondent's tendency to provide correct answers to the domain of questions that contribute to a scale as a function of a parameter called performance, estimated by a scale score. The scale scores can be viewed as a summary measure of performance across the domain of questions that make up the scale. Three distinct IRT models were used for scaling: three-parameter logistic models for multiple-choice questions; two-parameter logistic models for short constructed-response questions that were scored correct or incorrect; and generalized partial credit models for short and extended constructed-response questions that were scored on a multipoint scale (i.e., greater than two levels).

Three distinct scales were created for the national assessment program in science to summarize fourth-grade students' abilities according to the three defined fields of science (earth, physical, and life). Within each scale, the estimates of the empirical item characteristic functions were compared with the theoretical curves to determine how well the IRT model fit the observed data. For correct-incorrect questions, nonmodel-based estimates of the expected proportions of correct responses to each question for students with various levels of scale proficiency were compared with the fitted item response curve. For the short and extended partial-credit constructed-response questions, the comparisons were based on the expected proportions of students with various levels of scale proficiency who achieved each score level. In general, the scaling models fit the question-level results well.

Using the item parameter estimates from the national grade 4 assessment in science, estimates of various population statistics were obtained for DDESS and DoDDS. The NAEP methods use random draws ("plausible values") from estimated proficiency distributions for each student to compute population statistics. Plausible values are not optimal estimates of individual student proficiencies; instead, they serve as intermediate values to be used in estimating population characteristics. Under the assumptions of the scaling models, these population estimates will be consistent, in the sense that the estimates approach the model-based population values as the sample size increases, which would not be the case for population estimates obtained by aggregating optimal estimates of individual performance.

The 1996 science assessment was developed using a new framework. Because it was not appropriate to compare results from the 1996 assessment to those of previous NAEP science assessments, no attempt was made to link or align scores on the new assessment to those of previous assessments. Therefore, it was necessary to establish a new scale for reporting. Earlier NAEP assessments (such as the current mathematics assessment and the 1994 reading assessment) were developed with a cross-grade framework, in which the trait being measured is conceptualized as cumulative across the grades of the assessment. This concept was reflected in the scaling. The score scales developed for these assessments were cross-grade scales on a single 0-500 scale for all three grades in the assessment.

In 1993, the National Assessment Governing Board (NAGB) determined that future NAEP assessments should be developed using within-grade frameworks. This removes the constraint that the trait being measured is cumulative, and there is no need for overlap of questions across grades. Consistent with this view, NAGB also declared that scaling be performed within-grade. Any items which happened to be the same across grades in the assessment were scaled separately for each grade, thus allowing common items, potentially, to function differently in the separate grades. The 1994 NAEP history and geography assessments were developed and scaled within-grade. After scaling, the scales were aligned so that grade 8 had a higher mean than did grade 4, and grade 12 had a higher mean than grade 8. The results were reported on a final 0-500 scale that looked similar to those used in mathematics and reading, despite the differences in development and scaling. This definition of the reporting scale was a source of potential confusion and misinterpretation.

The 1996 science assessment was also developed and scaled using within-grade procedures. A new reporting metric was adopted to differ from the 0-to-500 reporting scales used in other NAEP subject areas in order to minimize confusion with other common test scales and to discourage inappropriate cross-grade comparisons. For each grade in the national assessment, the mean for each field of science was set at 150 and the standard deviation was set at 35. First, the reporting metric was developed using data from the national assessment program; the results for the DoDEA science assessment were then linked to that scale using procedures described in Section C.10.

In addition to the plausible values for each scale, a composite of the three fields of science scales was created as a measure of overall science performance; as for the individual fields of science scales, the mean of the composite scale was set to 150 with a standard deviation of 35.[12] This composite was a weighted average of the plausible values for the three fields of science scales. The scales were weighted proportionally to the relative importance assigned to each field of science in the science framework (see Table B.1). The definition of the composite scale for the DoDEA assessments was identical to that used for the national fourth-grade science assessments.

C.10 Scaling Procedures to Link DoDEA Results to the National Results

Because there was no 1996 fourth-grade state assessment in science, the assessment in DoDEA schools at this grade level required special data analysis and scaling procedures. The five steps in linking the state assessment results to the national results were modified to the following three:

- conventional item analysis;
- estimation of proficiency distributions based on the "plausible values" methodology; and
- creation of science composite plausible values.

All analyses were performed treating the DDESS and DoDDS schools as two separate jurisdictions. IRT item statistics from the national grade 4 science analysis were used directly in the analysis and their use precluded having to link the DoDEA scales to the national science scales. The use of national item parameters was necessary because there was no fourth-grade

[12] The national average of students in public and nonpublic schools combined is 150. The national average seen in the tables in this report is based on the average for public schools only (148).

state assessment and because the two DoDEA samples are not large enough for an independent IRT estimation of item parameters, such as was done for the grade 8 state sample.

Following standard practice in NAEP analyses, the item analyses were carried out in order to check the data. Item statistics were compared to those from the national fourth-grade assessment results, and no data problems were detected.

Using student item responses, data from the background questionnaires (student, teacher, and school) and national item parameters, conditioning model parameters were estimated using the CGROUP computer program, separately for the DDESS and the DoDDS samples.

These plausible values were transformed to the final science scales using the same transformation used with the national fourth-grade plausible values. For each scale, the linear transformation obtained for the national grade 4 science scale was of the form:

$$Y^* = k_1 + k_2 Y$$

where

Y = a scale score level in terms of the system of units of the provisional scale of the national assessment scaling (or a DoDEA scale score level)

Y^* = a scale score level in terms of the system of units comparable to those used for reporting the 1996 national science results

k_2 = 35 / (Original National Standard Deviation)

k_1 = 150.0 - k_2 [Original National Mean]

The constants for the three scales are displayed in Table C.1.

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE C.1

Transformation Constants: Grade 4 National to DoDEA Results

Fields of Science Scales	k_1	k_2
Earth Science	150.6685	34.0920
Physical Science	151.1681	34.9092
Life Science	150.5101	35.0857

The composite scale plausible values were computed as the arithmetic mean of the plausible values on the three scales. This is in accord with the framework specification that each field of science content area have approximately equal weight in the grade 4 instrument. The plausible values for all scales were then placed on the database for further analysis. Scale score means for various subgroups were computed from the results.

APPENDIX D

Teacher Preparation

Because teachers are key to improving science education, their background and professional development should be examined. Fourth-grade science teachers completed questionnaires about their background and training, including their experience, certification, undergraduate and graduate course work in science, and involvement in preservice education.

Consistent with procedures used throughout this report, the student was the unit of analysis. That is, the science teachers' responses were linked to their students, and the data reported are the *percentages of students taught by these teachers* rather than the *percentages of teachers*.

The tables in Appendix D represent only a few of the questions in the teacher questionnaire, and this small selection can give only a sketchy profile of the DoDEA teachers. A report scheduled to appear in early 1998 will explore more of the questions related to school and classroom policy and practices, to give a better picture of the nation's teachers[1].

[1] The interested reader can obtain additional information on teachers' characteristics and qualifications and the conditions under which they teach in *SASS by State* (NCES 96-312) from the 1993-94 Schools and Staffing Survey.
URL: http://www.ed.gov/NCES/pubs/96312.html

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE D.1

Grade 4 Teachers' Reports on their Highest Level of Education

What is the highest academic degree you hold?	Percentage	
	DDESS	Nation
Bachelor's degree	40 (1.3)	57 (3.0)
Master's degree	47 (1.2)	36 (2.8)
Education specialist's or professional diploma	12 (0.5)	6 (1.0)
Doctorate or professional degree	1 (0.1)	0 (****)

The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). **** Standard error estimates cannot be accurately determined
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE D.2

Grade 4 Teachers' Reports on their Major Fields of Study

What were your major fields of study? (multiple responses possible)	Percentage	
	DDESS	Nation
Undergraduate		
Education	41 (1.3)	38 (3.5)
Elementary education	86 (1.0)	78 (3.1)
Secondary education	10 (0.4)	4 (0.9)
Science education	8 (0.4)	6 (1.1)
Life science	2 (0.2)	4 (1.0)
Physical science	0 (****)	3 (0.8)
Earth science	1 (0.2)	2 (0.8)
Other	25 (1.3)	36 (3.0)
Graduate		
Education	24 (0.8)	30 (3.4)
Elementary education	51 (1.4)	48 (3.4)
Secondary education	1 (0.1)	1 (0.4)
Science education	6 (0.4)	5 (1.3)
Life science	1 (0.1)	2 (0.7)
Physical science	1 (0.1)	2 (0.6)
Earth science	1 (0.2)	1 (0.6)
Other	11 (0.9)	19 (2.5)
No graduate study	15 (1.1)	18 (2.5)

The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). **** Standard error estimates cannot be accurately determined

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

BEST COPY AVAILABLE

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE D.3

Grade 4 Teachers' Reports on their Teaching Certification

	Percentage	
	DDESS	Nation
What type of teaching certification do you have in this state in your main assignment field?		
I don't have a certificate in my main assignment field.	0 (****)	0 (****)
Certification by an accreditation body other than the state	3 (0.4)	0 (****)
Temporary, provisional, or emergency state certificate	2 (0.2)	3 (1.1)
Probationary state certificate (initial certificate)	1 (0.3)	2 (0.8)
Regular or standard state certificate	63 (1.4)	77 (2.2)
Advanced professional certificate	31 (1.5)	18 (2.1)
Do you have teaching certification in any of the following areas that is recognized by the state in which you teach? (multiple responses possible)		
Elementary or middle/junior high school education	93 (0.6)	97 (1.0)
Elementary science	36 (1.4)	43 (3.5)
Middle/junior high school or secondary science	13 (0.8)	18 (3.0)
Other	45 (2.2)	39 (4.3)

The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). **** Standard error estimates cannot be accurately determined.

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

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE D.4

Grade 4 Teachers' Reports on Years Teaching Experience

Counting this year, how many years have you . . .	Percentage	
	DDESS	Nation
taught at either the elementary or secondary level? [1]		
2 years or less	5 (0.8)	9 (1.3)
3-5 years	13 (1.0)	13 (1.6)
6-10 years	28 (1.4)	21 (2.2)
11-24 years	37 (1.3)	31 (2.7)
25 years or more	16 (0.7)	26 (2.7)
taught science? [2]		
2 years or less	9 (0.9)	12 (1.5)
3-5 years	24 (1.0)	16 (1.6)
6-10 years	21 (1.3)	21 (2.1)
11-24 years	36 (1.3)	32 (2.4)
25 years or more	10 (0.7)	19 (2.3)

The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details). [1]Teachers were instructed to include part-time teaching experience. [2] Teachers were instructed to include full-time and part-time assignments, but not substitute assignments.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE D.5

Grade 4 Teachers' Reports on Recent Course Taking

During the last two years, how many college or university courses have you taken in science or science education?	Percentage	
	DDESS	Nation
None	88 (0.9)	78 (3.0)
One	9 (0.8)	17 (2.8)
Two	1 (0.2)	3 (0.9)
Three or more	3 (0.3)	2 (0.8)

The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE D.6

Grade 4 Teachers' Reports on Professional Development Activities

	Percentage	
	DDESS	Nation
During the past two years, have you taken college or university courses in any of the following?		
Methods of teaching science	19 (0.8)	17 (2.0)
Biology/life science	13 (0.9)	10 (1.6)
Chemistry	6 (0.3)	5 (1.1)
Physics	4 (0.2)	4 (1.0)
Earth science	9 (0.6)	8 (1.6)
During the past five years, have you taken courses or participated in professional development activities in any of the following?		
Use of computers for data acquisition	38 (1.2)	33 (2.9)
Use of computers for data analysis	39 (1.3)	36 (2.8)
Use of multimedia for science education	46 (1.3)	33 (3.5)
Laboratory management or safety	8 (0.9)	9 (1.7)
Integrated science instruction	34 (1.1)	31 (2.9)

The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE D.7

Grade 4 Public School Teachers' Reports on Professional Development

During the last year, how much time in total have you spent in professional development workshops or seminars in science or science education?	Percentage	
	DDESS	Nation
None	29 (0.9)	31 (2.8)
Less than six hours	45 (1.2)	30 (2.6)
6-15 hours	19 (0.7)	23 (3.0)
16-35 hours	4 (0.4)	9 (1.6)
More than 35 hours	2 (0.2)	8 (2.1)

The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

THE NATION'S REPORT CARD naep 1996 State Assessment

TABLE D.8

Grade 4 Teachers' Reports on Membership in Professional Societies

Do you belong to one or more professional organizations related to science?	Percentage	
	DDESS	Nation
Yes	25 (1.0)	9 (1.3)
No	75 (1.0)	91 (1.3)

The standard errors of the statistics appear in parentheses. It can be said with about 95 percent confidence that, for each population of interest, the value for the entire population is within ± standard errors of the estimate for the sample. In comparing two estimates, one must use the standard error of the difference (see Appendix A for details).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

ACKNOWLEDGMENTS

This report is the culmination of the effort of many individuals who contributed their considerable knowledge, experience, and creativity to the NAEP 1996 science assessment. The NAEP 1996 science assessment was a collaborative effort among staff from the National Center for Education Statistics (NCES), the National Assessment Governing Board (NAGB), Educational Testing Service (ETS), Westat, Inc., and National Computer Systems (NCS). In addition, the program benefited from the contributions of hundreds of individuals at the state and local levels — governors, chief state school officers, state and district test directors, state coordinators, and district administrators — who tirelessly provided their wisdom, experience, and hard work. Most importantly, NAEP is grateful to the students, teachers, and administrators who made the completion of the science assessment possible.

The NAEP 1996 science assessment was funded through NCES, in the Office of Educational Research and Improvement of the U.S. Department of Education. The Commissioner of Education Statistics, Pascal D. Forgione, and the NCES staff — Janis Brown, Peggy Carr, Arnold Goldstein, and Steven Gorman— worked collegially with the authors to produce this report. In particular, the authors are indebted to Arnold Goldstein of NCES for coordinating the activities of the many people who contributed to this report.

The NAEP project at ETS is housed in the Center for the Assessment of Educational Progress under the direction of Paul Williams. The NAEP 1996 assessments were directed by Stephen Lazer and John Mazzeo. Tom Corley, Lee Jones, Tim Ligget, Beth Nichols, Christine O'Sullivan, Amy Pearlmutter, Will Pfeiffenberger, Mario Yepes-Baraya, and Ann Marie Zolandz worked with the Science Instrument Development committee to develop the assessment instrument. Sampling and data collection activities were conducted by Westat under the direction of Rene Slobasky, Nancy Caldwell, Keith Rust, and Dianne Walsh. Printing, distribution, scoring, and processing activities were conducted by NCS under the direction of Brad Thayer, Patrick Bourgeacq, Charles Brungardt, Jay Happel, Mathilde Kennel, Linda Reynolds, and Brent Studer.

The complex statistical and psychometric activities necessary to report results for the NAEP 1996 Science Assessment were directed by Nancy Allen, John Barone, James Carlson, and Juliet Shaffer. The analyses presented in this report were led by John Donoghue and Steven Isham with assistance from Spencer Swinton, Lois Worthington, Inge Novatkoski, Kate Pashley, David Freund, and Norma Norris.

Kellie Keiser checked the data and tables. Debbie Kline coordinated the technical appendices, and considerable additional input for the DoDEA grade 4 reports was provided by Jim Carlson and John Donoghue. The overall production efforts were completed by Karen Damiano, Carol Errickson, and Loretta Casalaina. The World Wide Web version of the state reports was produced by Philip Leung and Pat O'Reilly with assistance from colleagues.

Many thanks are due to the numerous reviewers, both internal and external to NCES and ETS. The comments and critical feedback of the following reviewers are reflected in this report and the science state report on which it is based: Sue Ahmed, Peggy Carr, Mary Frase, Arnold Goldstein, Andrew Kolstad, Michael Ross, and Shi-Chang Wu of NCES; Rolf Blank of CCSSO; Audrey Champagne of the State University of New York in Albany; Michelle Leon of the Connecticut Department of Education; Will Pfeiffenberger of ETS; Senta Raizen of the National Center for Improving Science Education; and Mistilina Sato of Stanford University.

NAEP 1996 Science Instrument Development Committee

An Instrument Development Committee was convened to oversee the development of items and scoring rubrics. Committee members wrote assessment exercises and ensured that the instrument adhered to the assessment framework and specifications. In addition, the committee made certain that the instrument was developmentally appropriate for each grade and that it was relevant to curricular and instructional goals. The members are to be commended for their diligence and dedication to the lengthy process of producing the instrument:

Gail Baxter, University of Michigan
Ron Bonnstetter, University of Nebraska
Audrey Champagne, State University of New York at Albany
Richard Clark, Minnetonka, Minnesota
Sally Crissman, Shady Hill School, Cambridge, Massachusetts
Pat Dung, Los Angeles Educational Partnership
Michael Johnson, Science Skills Center
Michael Jojola, Isleta, New Mexico
Clifton Poodry, University of California at Santa Cruz
Senta Raizen, National Center for Improving Science Education
Douglas Reynolds, Rensselaer, New York
Realista Rodriguez, Stuart High School, Falls Church, Virginia
Mistilina Sato, Stanford University
Gerald Weaver, University City High School, Philadelphia, Pennsylvania
Mary Louise Bellamy, National Association of Biology Teachers

NCES 98–486