NATIONAL CENTER FOR EDUCATION STATISTICS

Assessing the Best:

NAEP's 1996 Assessment of Twelfth-Graders Taking Advanced Science Courses

THE NATION'S REPORT CARD naep

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

NCES 2001–451

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

Assessing the Best:
NAEP's 1996 Assessment of Twelfth-Graders Taking Advanced Science Courses

Christine Y. O'Sullivan
Wendy S. Grigg

In collaboration with

Jay R. Campbell
Steve Isham
Jinming Zhang

August 2001

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

NCES 2001–451

U.S. Department of Education
Rod Paige
Secretary

Office of Educational Research and Improvement
Grover J. Whitehurst
Assistant Secretary

National Center for Education Statistics
Gary W. Phillips
Acting Commissioner

August 2001

SUGGESTED CITATION
U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics. *Assessing the Best: NAEP's 1996 Assessment of Twelfth-Graders Taking Advanced Science Courses,* NCES 2001–451, by C. Y. O'Sullivan & W. S. Grigg. Washington, DC: 2001.

FOR MORE INFORMATION
Content contact:
Holly Spurlock
202–502–7458

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

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

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

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

Due to the confidential nature of NAEP surveys, the photographs throughout this report do not portray actual students who participated in the NAEP science assessment. The photographs used are from Comstock, Eyewire Images, The Stock Market and PhotoDisc stock libraries.

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.

Table of Contents

Highlights

The National Assessment of Educational Progress (NAEP) is the nation's only ongoing survey of what students know and can do in various subject areas. Authorized by Congress and administered by the National Center for Education Statistics in the Department of Education, NAEP regularly reports to the public on the educational progress of students in grades 4, 8, and 12.

In addition to the main NAEP science assessment that was conducted at all three grade levels in 1996, a special study was done that focused on twelfth-grade students taking advanced science courses in biology, chemistry, or physics during the 1995–96 school year. The purpose of the study was to assess what the top science students in the country know and can do in these subject areas.

The results of the study are presented in this report, which includes information on the science courses students reported taking, their overall performance on the assessment, and performance results for selected questions. Students' overall performance on the advanced science assessment is reported using two scales, a biology scale and a chemistry/physics scale.[1] Wherever possible, information is also provided for students who participated in the 1996 main NAEP science assessment, including data for the subgroup of students who were not enrolled in advanced science courses.

The following are some of the major findings from this study:

- An estimated 23 percent of all twelfth-grade students were taking advanced science courses in the 1995–96 school year.
- Sixty-nine percent of students in the advanced science study and 23 percent of the students from the main NAEP assessment who were not enrolled in an advanced science course reported taking seven or more semesters of science.
- Female students who participated in the advanced science study were more likely than males to go beyond one year of course work in biology.

[1] The results for chemistry and physics were combined into a single scale in order to be consistent with the main NAEP science assessment in which similar questions were grouped together under the broad domain of "physical science."

- More than two-thirds of the students who participated in the advanced science study reported taking one or more years of biology (98 percent), chemistry (94 percent), or physics (70 percent). While a similar proportion of students who were not taking an advanced science course reported taking one or more years of biology (92 percent), there were fewer students taking one or more years of chemistry or physics (60 percent and 23 percent, respectively).
- Males outperformed females on questions that measured students' knowledge of chemistry and physics.
- White students and Asian/Pacific Islander students had higher scale scores than black students and Hispanic students for biology and chemistry/physics.
- The average scale scores of students in the advanced science study who attended public and nonpublic schools were about the same.
- Students in the advanced science study were more likely than the students in the national sample to respond correctly to the set of common questions administered to both groups. The difference in question scores between the advanced study and main NAEP samples on common questions ranged from 2 to 19 percentage points.
- In general, constructed-response questions in the advanced science study were more difficult than multiple-choice questions and tended to have a higher percentage of omits than multiple-choice questions. This was also true for the main NAEP assessment.

Chapter 1

Overview of the 1996 Advanced Science Study

Introduction

The *Goals 2000: Educate America Act of 1994* made improving student achievement in science a national priority by stating as one of its goals that "all students will leave grades 4, 8, and 12 having demonstrated competency over challenging subject matter."[1] As the nation's only ongoing survey of what students know and can do in various academic subjects, including science, the National Assessment of Educational Progress (NAEP) is a valuable tool for monitoring the country's progress in meeting this goal.

The framework of the 1996 main NAEP assessment in science was organized along two dimensions. The first dimension divided science into three major fields: earth, physical, and life sciences. The second dimension defined characteristic ways of knowing and doing science: conceptual understanding, scientific investigation, and practical reasoning. The recommendation to study twelfth-grade students taking advanced science courses came in response to criticism that the main NAEP assessment did not include an adequate number of questions at advanced levels of difficulty that could reflect what the best-prepared students know or can do in science.[2] The advanced science study was therefore designed to obtain information specifically related to the performance of the nation's top science students.

This chapter provides an overview of the study, including the nature of the samples of students whose performance is described in the report, the structure of the assessment, and how the results were analyzed. Chapter 2 provides information related to the science courses students reported taking and summarizes their performance on the advanced science assessment. Chapter 3 presents selected questions from the study along with samples of students' responses and question-level results.

This report also includes three appendices that augment the information presented in these chapters. Appendix A provides the scoring guides corresponding to the questions discussed in chapter 3. Appendix B contains a detailed description of the procedural aspects

[1] Goals 2000: Educate America Act of 1994, H.R. 1804, 103rd Cong., 2nd Sess. (1994).

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

of the advanced science study. The standard errors for all the data presented in chapters 2 and 3 are provided in appendix C, and are useful in determining the level of uncertainty associated with the estimated results (see pages 67–69 for a more detailed explanation on interpreting standard errors).

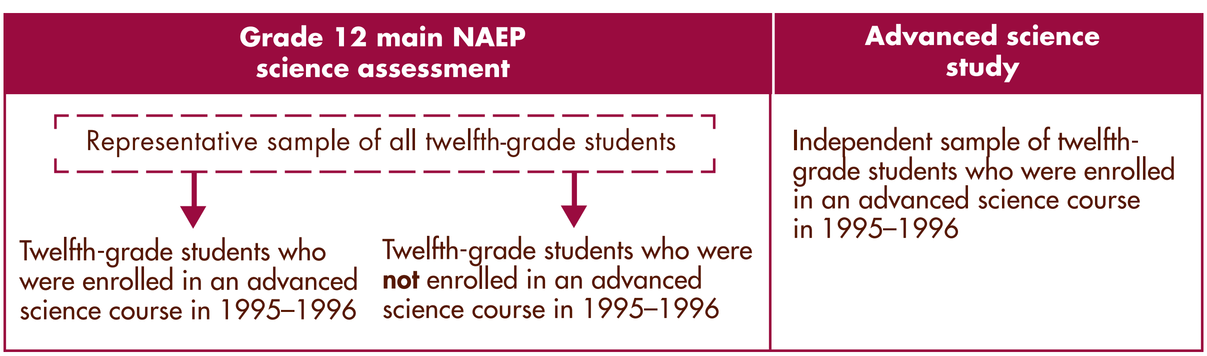
Student Samples

The 2,431 students who participated in the advanced science study represented an estimated 23 percent of all twelfth-grade students who were taking advanced-level science courses in the 1995–96 school year.[3] While the students who took part in the main NAEP assessment represented a cross section of different academic backgrounds, eligibility for the advanced science study was limited to those students who, according to school records, were enrolled in one or more of the following advanced science courses during the 1995–1996 school year: Advanced Placement (AP) Biology, Chemistry 2 or AP Chemistry, Physics 1, Physics 2 without calculus, or AP Physics. These courses were selected based on their perceived availability, their consistency in course work across schools, and because their subject domains were closely related to the existing NAEP science framework.

Figure 1.1 illustrates the sampling structure for the advanced science study and the main NAEP science assessment. In addition to the results obtained from the advanced science study sample, data are also presented in chapter 2 for the subset of students who took the 1996 main NAEP science assessment but were not eligible for the advanced science study (i.e., were not enrolled in an advanced science course). This allowed comparisons to be made between two distinctly different groups of students vis-à-vis science course-taking patterns. A more detailed description of the sampling process is included in appendix B.

Figure 1.1 ***Sampling structure for the grade 12 main NAEP science assessment and advanced science study: 1996***

THE NATION'S REPORT CARD naep



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

[3] The estimated 23 percent is based on school reported data collected during the 1996 main NAEP science assessment.

The Advanced Science Study Assessment

The questions developed for the advanced science study were deemed by a committee of science educators to reflect more challenging material than the main NAEP science assessment.[4] In the main assessment, the questions were based on the *Science Framework for the 1996 National Assessment of Educational Progress*[5] and measured knowledge and understanding in the three traditional fields of science: life science, earth science, and physical science. In the advanced science study, the questions measured students' knowledge of biology, chemistry, and physics.

Students participating in the study were given two hours to respond to a total of 66 questions—22 in the field of biology, 22 in chemistry, and 22 in physics. The questions were presented in several different formats including 30 multiple choice, 24 short constructed response (requiring a one- or two-sentence answer), and 12 extended constructed response (requiring a more in-depth answer of a paragraph or more). Table 1.1 provides a breakdown of the question format by domain. Eighteen questions from the main NAEP assessment were included among the 66 questions administered as part of the study to allow for some comparisons in students' performance between the two assessments. The remaining 48 questions were unique to the advanced science study (16 questions in each of the three fields of science assessed).

TABLE 1.1 ***Distribution of questions by field of science in the grade 12 NAEP advanced science study: 1996***

THE NATION'S REPORT CARD naep

Field of science	Multiple choice	Short constructed response	Extended constructed response	Total
Biology	10	10	2	22
Chemistry	10	8	4	22
Physics	10	6	6	22
Total	**30**	**24**	**12**	**66**

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

Reporting NAEP Results

Students' performance on the advanced science study assessment is presented in this report both in terms of overall performance and performance on individual questions. As with the main NAEP science assessment, students' overall performance on the advanced science assessment is reported using average scale scores that range from 0 to 300. Whereas individual scales were created for each of the three fields of science assessed in the main assessment (life,

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

[5] Ibid.

physical, and earth), the results from the advanced science study were reported using two scales: a biology scale and a chemistry/physics scale. These two scales were created by linking the results for students in the advanced science study to the performance of a comparable subgroup of students who took the main assessment even though they were eligible to participate in the advanced science study. Results from the biology questions on the advanced science assessment were linked to the life science scale from the main assessment, and the chemistry and physics results were combined so that they could be linked to the physical science scale. There were no questions in the advanced science study that were comparable to the earth science questions in the main assessment. A detailed explanation of the linking process is included in appendix B.

While the scale score ranges were identical, each content area scale was derived independently. Therefore, average scale scores across the two content areas (biology and chemistry/physics) cannot be compared to each other or to the grade 12 composite scale developed for the main assessment. For example, equal scores on the biology and chemistry/physics scales do not imply equal levels of science achievement in these domains. Further details on the scaling methodology can be found in *The NAEP 1996 Technical Report*.[6]

Results pertaining to students' performance on individual assessment questions are reported as the percentage of students choosing each response option for a multiple-choice question and the percentage achieving each of the score categories for a constructed-response question. Each constructed-response question had a unique scoring guide that specified the range of possible scores and defined the criteria used to evaluate students' responses. Short constructed-response questions were scored according to three levels of performance, *Complete*, *Partial*, or *Unsatisfactory*. Extended constructed-response questions were scored using four levels: *Complete*, *Essential*, *Partial*, or *Unsatisfactory*.

In addition to providing information on how students responded to individual sample questions, some questions have been mapped on the 0 to 300 scale in order to provide a visual representation of their difficulty in relation to other questions and to students' overall performance on the assessment. A detailed explanation of how to interpret this information and the corresponding item map are on pages 40 and 41.

Cautions in Interpretations

The reader is cautioned against using the NAEP results presented in this report to make simple causal inferences related to student performance or to the effectiveness of public and nonpublic schools. A relationship that exists between performance and another variable does not reveal its underlying cause. While this report focuses on course-taking, there may be other variables that could contribute to students' performance. For example, differences in science performance may reflect a range of socioeconomic and educational factors not discussed in this report or addressed by the NAEP program.

---

[6] Allen, N. L., Carlson, J., & Zelenak, C. A. (1999). *The NAEP 1996 technical report*. Washington, DC: National Center for Education Statistics.

Chapter 2

Science Course-Taking and Student Performance

Introduction

One of the most important, though not surprising, findings of the 1996 main NAEP science assessment was that the more science courses students took, the better they performed on the assessment.[1] This should be tempered with the observation that physics appeared to be a critical course; students who took any combination of subjects that included physics outperformed their peers who did not take physics as one of their courses.[2]

Students in the advanced science study were asked the same background questions as those students who took the main NAEP assessment. The first section of this chapter describes course-taking patterns reported by students in the advanced science study and students in the main assessment who were not eligible for the study (i.e., were not enrolled in an advanced science course). The second section of this chapter presents information on the performance of students in the advanced science study. The data presented in this chapter reflect both public and nonpublic schools combined, unless stated otherwise.

[1] O'Sullivan, C. Y., Weiss, A. R., & Askew, J. M. (1998). *Students learning science: A report on policies and practices in U.S. schools*. (NCES Publication No. 98–493). Washington, DC: National Center for Education Statistics.

[2] Ibid.

Number of Semesters of Science Taken, Grades 9 –12

Most school districts require two or three years of science course work for graduation;[3] however, many students choose to take more courses. Table 2.1 shows the reported number of semesters of science taken by students in the advanced science study and those taking the main NAEP science assessment who were not eligible to participate in the advanced study. Sixty-nine percent of students who participated in the advanced science study reported having taken seven or more semesters of science. A higher percentage of white students than Hispanic students who took the advanced science assessment reported taking seven or more semesters of science. Although there also appears to be a difference between white students and black students taking seven or more semesters of science, it was not found to be significant.

Table 2.1 ***Students' reports on number of semesters of science taken from grades 9–12, by gender and race/ethnicity: 1996***

THE NATION'S REPORT CARD naep

From the beginning of ninth grade to the present, how many semesters of course work will you have taken in science?	Percentage of students				
	7 or more semesters	5-6 semesters	3-4 semesters	1-2 semesters	No semesters
Total					
Advanced study	69	16	15	1	0
Not eligible*	23	29	36	10	2
Male					
Advanced study	68	17	14	1	0
Not eligible*	23	28	35	12	2
Female					
Advanced study	69	15	15	1	0
Not eligible*	23	30	38	9	1
White					
Advanced study	72	15	13	0	0
Not eligible*	25	31	35	8	1
Black					
Advanced study	58	13	25	3	0
Not eligible*	18	21	43	15	3
Hispanic					
Advanced study	49	24	24	1	1
Not eligible*	18	25	38	17	3
Asian/Pacific Islander					
Advanced study	66	20	12	2	0
Not eligible*	27	29	35	9	1

*"Not eligible" refers to students who participated in the main NAEP assessment but were not enrolled in an advanced science course.

NOTE: Percentages may not add to 100 due to rounding.

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

[3] O'Sullivan, C. Y., Weiss, A. R., & Askew, J. M. (1998). *Students learning science: A report on policies and practices in U.S. schools*. (NCES Publication No. 98-493). Washington, DC: National Center for Education Statistics.

Amount of Course Work Taken in Biology

Based on the information obtained from the main NAEP assessment, 96 percent of all students in grade 12 reported completing at least some course work in biology.[4] Many students, however, opted to take a second year of biology. Table 2.2 indicates that almost all of the students who participated in the advanced science study and the non-eligible students from the main assessment reported taking at least one year of biology (98 percent and 92 percent, respectively). Thirty percent of students who participated in the advanced science study reported taking more than one year of biology. Female students who participated in the study were more likely than male students to go beyond one year of course work in biology.

Table 2.2 — ***Students' reports on amount of course work taken in biology from grades 9–12, by gender and race/ethnicity: 1996***

THE NATION'S REPORT CARD naep

From the beginning of ninth grade to the present, how much science course work have you completed in biology?	Percentage of students			
	> 1 year	**1 year**	**< 1 year**	**None**
Total				
Advanced study	30	68	1	1
Not eligible*	14	78	3	5
Male				
Advanced study	26	72	1	2
Not eligible*	15	76	3	6
Female				
Advanced study	34	64	1	1
Not eligible*	13	80	3	4
White				
Advanced study	32	66	1	1
Not eligible*	15	79	2	4
Black				
Advanced study	19	80	0	1
Not eligible*	13	79	3	5
Hispanic				
Advanced study	16	82	1	2
Not eligible*	12	73	6	8
Asian/Pacific Islander				
Advanced study	39	59	1	1
Not eligible*	10	80	5	5

*"Not eligible" refers to students who participated in the main NAEP assessment but were not enrolled in an advanced science course.

NOTE: Percentages may not add to 100 due to rounding.

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

---

[4] Ibid.

Amount of Course Work Taken in Chemistry

Since chemistry is often offered as a third year of science, students who have fulfilled their graduation requirements may not feel the need to take a course in chemistry. As a result, fewer students take a course in chemistry than biology. In twelfth grade, 74 percent of students who took the main NAEP assessment reported taking chemistry.[5] As shown in table 2.3, 94 percent of students who participated in the advanced science study had taken one or more years of chemistry. Sixty percent of students who were not eligible for the study indicated that they had taken one or more years of chemistry. Since students who were eligible to be in the advanced study had to be enrolled in at least one advanced science class in the 1995–96 school year, it is possible that the five percent of non-eligible students who reported taking more than one year of chemistry had taken Chemistry 2 or AP Chemistry prior to the 1995–96 school year.

Table 2.3 ***Students' reports on amount of course work taken in chemistry from grades 9–12, by gender and race/ethnicity: 1996***

THE NATION'S REPORT CARD naep

From the beginning of ninth grade to the present, how much science course work have you completed in chemistry?	Percentage of students			
	> 1 Year	1 Year	< 1 Year	None
Total				
Advanced study	16	78	2	4
Not eligible*	5	55	7	33
Male				
Advanced study	18	75	2	5
Not eligible*	6	52	6	35
Female				
Advanced study	14	81	2	4
Not eligible*	4	58	7	31
White				
Advanced study	16	79	2	4
Not eligible*	5	56	6	33
Black				
Advanced study	12	82	3	3
Not eligible*	7	54	7	32
Hispanic				
Advanced study	9	81	2	8
Not eligible*	5	50	8	38
Asian/Pacific Islander				
Advanced study	27	68	3	2
Not eligible*	6	64	7	23

*"Not eligible" refers to students who participated in the main NAEP assessment but were not enrolled in an advanced science course.

NOTE: Percentages may not add to 100 due to rounding.

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

[5] O'Sullivan, C.Y., Weiss, A.W., & Askew, J. M. (1998). *Students learning science; A report on policies and practices in U.S. schools* (NCES Publication No. 98–493). Washington, DC: National Center for Education Statistics.

Amount of Course Work Taken in Physics

Data from the 1996 main NAEP assessment show that 41 percent of twelfth-grade students had taken a course in physics.[6] The percentage of students taking physics may be lower than for biology or chemistry for a variety of reasons. For example, physics is traditionally offered as an eleventh- or twelfth-grade course, and students who have already fulfilled the science requirement for graduation may opt not to take another science course. Alternatively, students may not be able to enroll in a physics course because of prerequisite mathematics courses and, as a result, may choose to take a second year of biology or a course in science and technology instead. Table 2.4 presents the results pertaining to physics course-taking. While 70 percent of students participating in the advanced science study reported taking one year or more of physics, 20 percent reported taking no physics. More males than females in the advanced science study reported taking one year of physics. Twenty-three percent of students who were not eligible for the special study reported taking one or more years of physics, presumably prior to the 1995–96 school year. Seventy-one percent of the non-eligible students reported not having taken course work in physics.

---

[6] Ibid.

Table 2.4

Students' reports on amount of course work taken in physics from grades 9–12, by gender and race/ethnicity: 1996

From the beginning of ninth grade to the present, how much science course work have you completed in physics?	Percentage of students			
	> 1 Year	1 Year	< 1 Year	None
Total				
Advanced study	8	62	10	20
Not eligible*	3	20	5	71
Male				
Advanced study	10	66	8	16
Not eligible*	4	23	6	67
Female				
Advanced study	6	59	11	25
Not eligible*	2	18	5	75
White				
Advanced study	6	63	9	21
Not eligible*	3	21	5	71
Black				
Advanced study	5	61	9	24
Not eligible*	2	18	6	74
Hispanic				
Advanced study	6	62	11	21
Not eligible*	3	18	6	73
Asian/Pacific Islander				
Advanced study	20	59	11	10
Not eligible*	9	28	9	54

*"Not eligible" refers to students who participated in the main NAEP assessment but were not enrolled in an advanced science course.

NOTE: Percentages may not add to 100 due to rounding.

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

Advanced Placement Courses Taken

Advanced Placement (AP) courses are academically challenging. Students take them and the subsequent examinations in order to receive college credit or to gain competitive advantages in college admissions. As part of the main NAEP assessment and the advanced science study, students were asked to indicate whether they were currently enrolled in an AP course in science. It should be noted that this question referred to the current school year only and did not take previous years into account. Although students had to be enrolled in an advanced-level science course to be eligible to participate in the study, not all of the advanced courses were necessarily AP courses. For example, students could have been taking Chemistry 2 or Physics 1. Table 2.5 shows that less than half (43 percent) of the students in the advanced science study were enrolled in an AP course in science. Only 16 percent of the students participating in the main assessment reported being enrolled in an AP science course.[7]

Table 2.5 Students' reports on enrollment in Advanced Placement courses in science, by gender and race/ethnicity: 1996

THE NATION'S REPORT CARD naep

Are you currently enrolled in an Advanced Placement course in science?	Yes	No
Total Advanced study	43	57
Male Advanced study	44	56
Female Advanced study	42	58
White Advanced study	43	57
Black Advanced study	41	59
Hispanic Advanced study	28	72
Asian American/Pacific Islander Advanced study	57	43

NOTE: All students currently taking AP courses would have been eligible for the advanced science study—therefore, "Not eligible" results are not given.

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

[7] National Center for Education Statistics (1996). NAEP 1996 Summary Data Tables [Electronic data file]. Available: http://nces.ed.gov/nationsreportcard/tables96/nt1251.pdf

Student Performance on the Advanced Science Study

Performance of Selected Subgroups by Biology Scale and Chemistry/Physics Scale

The average scale scores for biology and chemistry/physics are presented in tables 2.6 and 2.7 for subgroups of students who participated in the advanced science study. Based on a scale of 0 to 300, the average scale scores were 173 for biology and 175 for chemistry/physics. The average scale scores for the main NAEP science assessment were 150 for life science and 150 for physical science. The advanced science study scales were created by linking the results for students in this study to the performance of a comparable subgroup of students who took the main assessment (i.e., those students, who although enrolled in an advanced science course, took the main assessment). Results from the biology questions on the advanced science assessment were linked to the life science scale from the main assessment, and the chemistry and physics results were combined so that they could be linked to the physical science scale. (A more detailed description of the linking process is included in appendix B.)

Gender. The performance of males and females on those questions that measured biology did not differ significantly, whereas males outperformed females for those questions that measured students' knowledge of chemistry and physics.

Race/Ethnicity. Both white students and Asian/Pacific Islander students had scale scores that were higher than those of black students and Hispanic students for both biology and chemistry/physics questions.

Type of School. Past NAEP results across a variety of subjects, including science, have consistently shown that students attending nonpublic schools outperform those attending public schools.[8] In the advanced science study, no significant differences were observed in the scale scores of students who attended public and nonpublic schools.

---

[8] O'Sullivan, C. Y., Reese, C. M., & Mazzeo, J. (1997). *NAEP 1996 science report card for the nation and the states: Findings from the National Assessment of Educational Progress* (NCES Publication No. 97–499). Washington DC: National Center for Education Statistics.

Table 2.6 Biology advanced science study scale scores, by gender, race/ethnicity, and type of school: 1996

THE NATION'S REPORT CARD naep

	Biology scale score
Total	**173**
Male	174
Female	172
White	178
Black	149
Hispanic	155
Asian/Pacific Islander	172
Public schools	173
Nonpublic schools	172
Catholic schools	170
Other nonpublic	175

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

Table 2.7 Chemistry/physics advanced science study scale scores, by gender, race/ethnicity, and type of school: 1996

THE NATION'S REPORT CARD naep

	Chemistry/physics scale score
Total	**175**
Male	181
Female	169
White	180
Black	150
Hispanic	153
Asian/Pacific Islander	178
Public schools	175
Nonpublic schools	172
Catholic schools	171
Other nonpublic	175

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

Summary

This chapter summarizes information about science course-taking patterns reported by students who participated in the advanced science study and those students in the main assessment who were not eligible for the study. It also provides the scale score results for the students in the advanced science study.

Course Work

- Sixty-nine percent of students in the advanced science study and 23 percent of the students from the main NAEP assessment who were not enrolled in an advanced science course reported taking seven or more semesters of science.
- Female students who participated in the advanced science study were more likely than males to go beyond one year of course work in biology.
- More than two-thirds of the students who participated in the advanced science study reported taking one or more years of biology (98 percent), chemistry (94 percent), or physics (70 percent). While a similar proportion of students who were not taking an advanced science course reported taking one or more years of biology (92 percent), there were fewer students taking one or more years of chemistry or physics (60 percent and 23 percent, respectively).

Performance

- Males outperformed females on those questions that measured students' knowledge of chemistry and physics.
- White students and Asian/Pacific Islander students had scale scores that were higher than those of black students and Hispanic students for both biology and chemistry/physics.
- No significant differences were observed in scale scores of students who attended public schools and students who attended nonpublic schools.

Chapter 3

Sample Questions and Student Responses

Introduction

This chapter contains sample questions for each of the major content area domains. Table 3.1 summarizes how each of the eleven questions presented in the chapter, two of which were common to both the advanced science study and the main NAEP assessment, were categorized by content area and question format. Since any detailed discussion of individual test questions must be limited to those questions that have been released to the public, there are no extended constructed-response questions in biology included in this section.

Table 3.1 **Sample questions categorized by content domain and question format**

THE NATION'S REPORT CARD naep

Type of question	Biology	Chemistry	Physics
Multiple choice	• Plant adaptation in desert	• Neutralization • Stoichiometry[1]	• Kinetic energy of block • Path of car on ice[1]
Short constructed response	• Function of vaccine • Genetic counseling	• Water at standard temperature and pressure	• Length of string
Extended constructed response		• Ionization energy	• Direction and speed of train relative to car

[1] Question was included in both the advanced science study and the main NAEP assessment.

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

Each sample question is presented along with a table containing the percentage of students choosing each response option for a multiple-choice question, or in the case of a constructed-response question, the percentage of students at each score level. In determining these percentages for the cognitive questions, a distinction was made between missing responses at the end of a block (i.e., missing responses following the last question the student answered) and missing responses prior to the last observed response. Missing responses before the last observed response were considered intentional omissions. Missing responses at the end of the block were considered "not reached" and treated as if the questions had not been presented to the student. In calculating response percentages for each question, only students classified as having been presented the question were included in the denominator of the statistic. Short constructed-response questions were scored on a three-point scale: *Complete*, *Partial*, or *Unsatisfactory*, while extended constructed-response questions were scored on a four-point scale: *Complete*, *Essential*, *Partial* or *Unsatisfactory*. Each constructed-response question had a unique scoring guide that identified the range of possible scores for the question and specified the criteria for evaluating student responses. The scoring guides corresponding to the sample constructed-response questions that follow are provided in appendix A of this report.

Biology

Twenty-two questions contributed to the biology subscale in the advanced science study, six of which were also included as part of the main NAEP assessment. One multiple-choice and two short constructed-response questions are described in this section.

Plant Adaptation in Desert

This multiple-choice question asked students to recognize structures characteristic of desert plants. To answer this question correctly, students had to know that leaves typically lose a lot of water and therefore tend to be reduced in size in desert plants. They also had to recognize that roots take in water and need to be deep in order to access water.

Which of the following are typical of plants that are adapted to desert conditions?

- (A) Broad, dark-green leaves and buttress roots
- (B) Narrow, thin leaves and reduced roots
- ● Reduced or nonexistent leaves and deep roots
- (D) Broad, thin leaves and taproots

Student performance data for this question are shown in table 3.2. Plant questions tend to be rather challenging. However, a large majority of students in the advanced science study answered this question correctly. Only a small percentage of students chose each of the three incorrect options.

Table 3.2 ***Plant adaptation in desert: Percentage choosing each response: 1996***

THE NATION'S REPORT CARD naep

A	B	C	D	Omit
6	8	80	6	0†

NOTE: Percentages may not add to 100 due to rounding.

†Percentage is between 0.0 and 0.5.

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

Function of Vaccine

The immune system is one of the body systems studied at the high school level. It is, however, fairly complex and tends to be covered somewhat superficially. The following short constructed-response question asks students to explain how vaccines work. The question was scored using a three-point scale. In order to receive a score of *Complete*, students needed to know that vaccines contain pieces of the infectious material (either modified or killed). They also needed to know that a vaccine causes an immune response; that is, cells are produced to fight off the infection. Some of these cells remain as memory cells that can respond immediately if the pathogen invades the body again. A score of *Partial* was given to responses that demonstrated some understanding of the process. For example, students may have known that vaccines cause the body to be prepared for another encounter with the disease-causing organism, but provided no further explanation. A score of *Unsatisfactory* was given to responses that showed minimal or no understanding of immunity.

> Children must be immunized against childhood diseases such as measles and mumps in order to attend public school. Explain how the vaccine used for one of these immunizations confers immunity.
>
> ______________________________
> ______________________________
> ______________________________
> ______________________________
> ______________________________

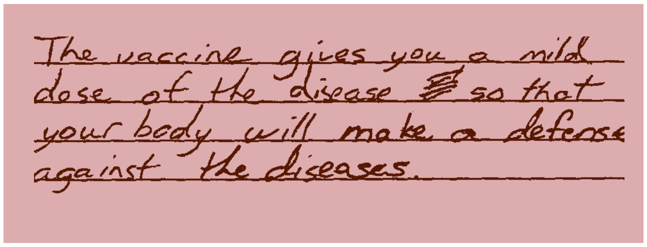
Sample 1: Function of Vaccine

The vaccine is a dead germ of the virus you want to be imunized against. You immune system creats antibodies to kill the already dead germ. The immune system's memory cell or (Killer-T-cell) remember how to create this specific antibody when the live virus acutally attacks the body.

Sample 1: *Complete* Response

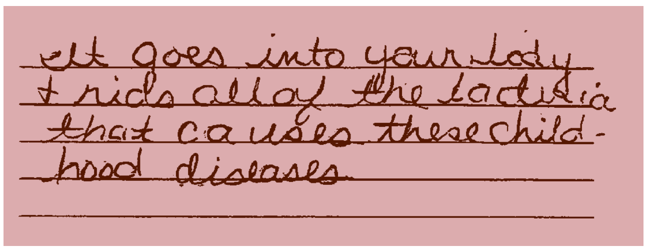
The first sample response received a score of *Complete*. The student stated that the vaccine "is a dead germ of the virus," and explained correctly that the memory cell "remembers how to create this specific antibody when the live virus actually attacks the body." The student demonstrated a fairly sophisticated understanding of the immune system by using the correct scientific terminology.

Sample 2: Function of Vaccine


The vaccine gives you a mild dose of the disease so that your body will make a defense against the diseases.

Sample 2: *Partial* Response
The second sample response received a score of *Partial*. This student knew that the vaccine gave a "mild dose of the disease," but failed to adequately explain how the body made "a defense against the disease."

Sample 3: Function of Vaccine


It goes into your body & rids all of the bactiria that causes these child-hood diseases

Sample 3: *Unsatisfactory* Response
The next sample shows a typical response that received a score of *Unsatisfactory*. This student clearly had no understanding of how vaccines work.

Table 3.3 shows the percentage of students at each of the score levels. About one-quarter of the students in the advanced science study were able to provide a complete response to this question. One-third of students were unable to explain how vaccines confer immunity and thus received a score of *Unsatisfactory*.

Table 3.3 ***Function of vaccine: Percentage at each score level: 1996***

THE NATION'S REPORT CARD naep

Unsatisfactory	*Partial*	*Complete*	Omit
33	34	23	8

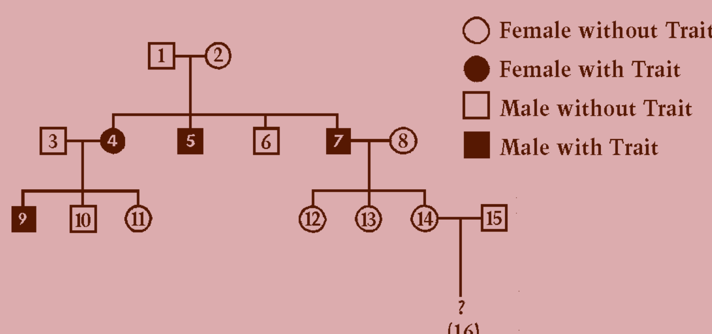
NOTE: Percentages may not add to 100 due to rounding.

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

Genetic Counseling

The purpose of the following short constructed-response question was to ascertain whether students could apply their knowledge of simple genetics to a real-life situation. As portrayed in the question, a young woman wished to know whether the child she is expecting is likely to be born with a genetic trait that is common in her family, but had not been observed in her husband's family. Students were presented with a pedigree and asked whether the child would have the trait and to explain their reasoning. In order to answer this question successfully, students had to have some understanding of dominant and recessive alleles and recognize that the trait in question resulted from two recessive alleles and was not linked to the XY chromosomes. They had to recognize that individuals [1] and (2) (as shown in the figure accompanying the question) both carried the trait since three of their four offspring ((4), [5], and [7]) had the trait. Thus the genotype of the parents must have been Tt, where T denotes the dominant allele and t denotes presence of the recessive allele. Similarly, the genotypes of offspring (4), [5], and [7] must have been tt, otherwise the trait would not have manifested itself. Because the young woman (number (14) in the pedigree) had a father with the trait (tt), she was most likely a carrier (Tt). Her husband was probably not a carrier (TT) since there was no evidence of the trait in his family. Thus their child (number 16) would have a 25 percent chance of being a carrier (Tt) and a 75 percent chance of not being a carrier (TT). Alternatively an argument could be made that the husband might carry the trait, in which case there would be a 25 percent chance of the child having the trait (tt), a 50 percent chance of carrying the trait (Tt), and a 25 percent chance of not carrying

A young woman who is pregnant seeks genetic counseling. She wants to know whether her child is likely to be born with a genetic trait that is common in her family, though it has not been observed in her husband's family. The counselor records the occurrence of the trait in her family and constructs the following pedigree. The woman is represented by 14 in the pedigree.



Will the woman's child (16 in the pedigree) have the trait? Explain the reasoning behind your conclusion.

____

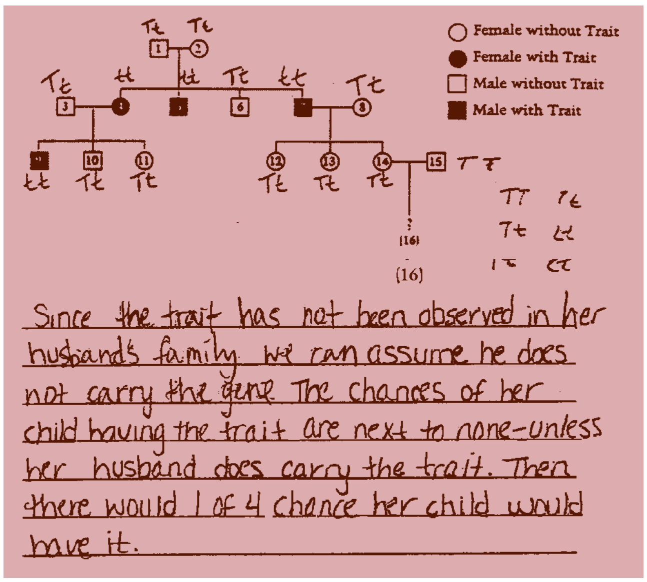
____

____

____

the trait (TT). Student responses were scored according to a three-point scale. Students whose responses were rated *Complete* were able to explain why the child was unlikely to carry the trait, or that the child had a chance of being a carrier or having the trait depending on whether or not the father carried the trait. To achieve a score of *Partial,* a student had to mention that the child is unlikely to show the trait, but may have been unclear in explaining why. A response that received a score of *Unsatisfactory* demonstrated minimal or no understanding of simple genetics.

Sample 1: Genetic Counseling

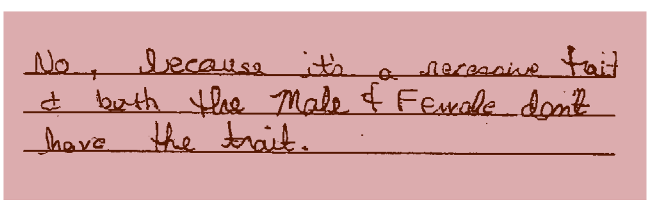


Since the trait has not been observed in her husband's family we can assume he does not carry the gene The chances of her child having the trait are next to none-unless her husband does carry the trait. Then there would 1 of 4 chance her child would have it.

Sample 1: *Complete* Response

The first sample response received a score of *Complete*. The student demonstrated a clear understanding of the pedigree as evidenced by the letters written by each member—using "T" to indicate dominant and "t" to indicate recessive. This student also considered whether or not the father might carry the trait and provided a reasonable explanation. To help bolster the argument for the second situation—the father being a carrier—the student included a Punnett square for a cross between two carriers and correctly stated that, if the father were a carrier, the child would have a 25 percent chance of having the trait.

Sample 2: Genetic Counseling

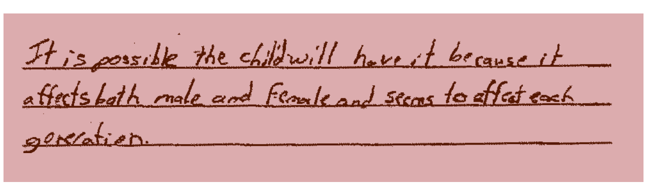


No, because it's a recessive trait & both the Male & Female don't have the trait.

Sample 2: *Partial* Response

The second sample response received a score of *Partial*. This student knew that the trait was recessive, but the reasoning was unclear.

Sample 3: Genetic Counseling



It is possible the child will have it because it affects both male and Female and seems to affect each generation.

Sample 3: *Unsatisfactory* Response

The next sample shows a typical response that received a score of *Unsatisfactory*. This student described the table, but did not relate it to dominant and recessive alleles and seemed to think that "It is possible the child will have it…."

Table 3.4 presents the percentage of students at each of the score levels. The question proved to be difficult, with only 8 percent of the students providing a reasoned discussion of a prediction they made based on a given pedigree. Twelve percent had a partial understanding of the genetics question, as evidenced by their ability to read the pedigree, but failed to give an adequate justification. Although the content of textbooks suggests that genetics is covered fairly substantially in high school biology courses, 75 percent of the students in the advanced science study received a score of *Unsatisfactory*.

Table 3.4 ***Genetic counseling: Percentage at each score level: 1996***

THE NATION'S REPORT CARD NAEP

Unsatisfactory	*Partial*	*Complete*	Omit
75	12	8	4

NOTE: Percentages may not add to 100 due to rounding.

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

Chemistry

Twenty-two questions made up the part of the advanced science study devoted to the field of chemistry. Six of these were included in the common block of questions from the main NAEP assessment, and the remaining 16 made up a separate section, which included a periodic table. Two multiple-choice questions (one of which was common to both the main assessment and advanced science study), one short constructed-response question, and one extended constructed-response question are described in this section.

Volume to Neutralize NaOH

Given the concentrations of an acid and a base, students were asked how much acid would neutralize 50 milliliters of base. The topic is fairly basic and is covered in first-year chemistry textbooks. In order to answer this question correctly, students had to recognize that HCl and NaOH dissociate into hydrogen and chloride ions and sodium and hydroxide ions. When the solutions are mixed, the hydrogen ions and the hydroxide ions react with each other to form water molecules. This reaction is referred to as "neutralization." Since the concentration of NaOH is twice that of HCl, it would require twice the volume of HCl solution to neutralize the NaOH. Thus the correct option is C.

How many milliliters of 0.10-molar HCl solution would be needed to completely neutralize 50 milliliters of 0.20-molar NaOH

Ⓐ 25 ml

Ⓑ 50 ml

● 100 ml

Ⓓ 200 ml

Information on student performance is presented in table 3.5. Sixty-five percent of the students in the advanced science study answered the question correctly. Options A and B proved to be fairly attractive, and students selecting these options most likely did not understand the mechanism of neutralization or the meaning of molar solutions.

Table 3.5 ***Volume to neutralize NaOH: Percentage choosing each response: 1996***

THE NATION'S REPORT CARD naep

A	B	C	D	Omit
15	14	65	4	2

NOTE: Percentages may not add to 100 due to rounding.

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

Grams Used in Reaction

This multiple-choice question was administered to students in the advanced science study and to students in the main NAEP assessment. In order to answer this question successfully, students had to use the mole ratios given by the coefficients of the balanced equation to relate those of one substance to those of another.

$$\text{i.e., } C + O_2 \longrightarrow CO_2$$

They then had to relate moles of one substance to mass in grams of that substance. The atomic mass of carbon is 12 and that of oxygen 16. Thus 12 grams of carbon combine with 32 grams of oxygen (diatomic) to form 44 grams of carbon dioxide (12 grams + 32 grams).

In an experiment, 12.0 grams of solid carbon reacted with oxygen gas to form 44.0 grams of carbon dioxide gas. How many grams of oxygen reacted with the carbon?

- (A) 12.0 grams
- ● 32.0 grams
- (C) 44.0 grams
- (D) 56.0 grams

Table 3.6 shows the percentage of students who participated in the advanced science study who chose each response option. Eighty-eight percent of the students were able to answer the question correctly. Students who responded by choosing options A, C, or D clearly did not understand basic stoichiometry.

Table 3.6 **Grams used in reaction: Percentage choosing each response: 1996**

THE NATION'S REPORT CARD naep

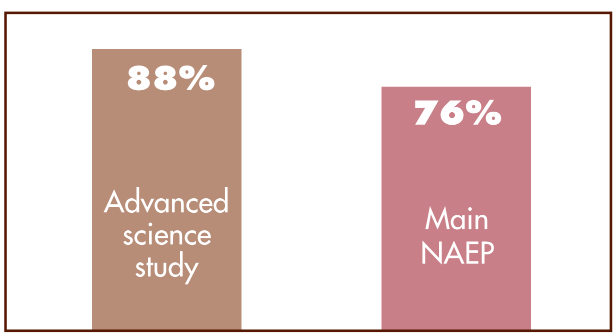
A	B	C	D	Omit
5	88	3	3	0†

NOTE: Percentages may not add to 100 due to rounding.
†Percentage is between 0.0 and 0.5.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Figure 3.1 shows the percentage of students in the advanced science study and the main NAEP assessment who chose the correct answer and suggests that students found the question fairly easy. Eighty-eight percent of students in the advanced science study chose the correct response. In the main NAEP, 76 percent of students answered correctly.

Figure 3.1 **Advanced science study and main NAEP — Grams used in reaction: Percentage choosing the correct response: 1996**

THE NATION'S REPORT CARD naep



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

Water at Standard Temperature and Pressure (STP)

The following short constructed-response question required students to explain why water is a liquid at room temperature and normal atmospheric pressure. They had to base their explanations on chemical bonding and intermolecular forces. This question was scored according to a three-point scale. In order to receive a score of *Complete*, students had to explain that water molecules are polar and that the bonds between O and H are polar bonds. Students had to realize that there is a net negative charge at the O atom and a net positive charge at the H atoms. Hydrogen bonds form between the O atom of one molecule and the H atom of another molecule. These bonds are strong and hold the molecules tightly together at room temperature. Students' responses that had at least some correct information received a score of *Partial*. A score of *Unsatisfactory* was given to responses that showed no indication that the students understood the concept.

Because of its low molecular mass, H_2O would be expected to be a gas at room temperature and normal atmospheric pressure. Explain in terms of chemical bonding and intermolecular forces why it is a liquid under these conditions and not a gas.

____

____

____

____

____

Sample 1: Water at STP

> H_2O is very polar so that the polar end one H_2O molecule is attracted to the negative end of another. These hydrogen bond a strong and so the temperature must be high to break down and make H_2O a gas.

Sample 1: *Complete* Response

The first sample response represents a full response and received a score of *Complete*. The student clearly knew that water molecules are polar. The student also knew that the bonds between water molecules are called "hydrogen bonds" and that they are strong and, hence, the temperature has to be high to break them.

Sample 2: Water at STP

> H_2O is a polar bond In a polar bond one side of the molecule is positi and the other side is negative

Sample 2: *Partial* Response

The second sample response received a score of *Partial*. This student knew that the water molecule is polar, but did not explain why this polarity leads to water being a liquid at standard temperature and pressure.

Sample 3: Water at STP

> When H and O bonds and becomes H_2O it preciptates into liquid

Sample 3: *Unsatisfactory* Response

The next sample shows a typical response that received a score of *Unsatisfactory*. This student clearly had no understanding of why water is a liquid at standard temperature and pressure and believed that, when hydrogen and oxygen bond, they precipitate to form a liquid.

Table 3.7 shows the percentage of students at each of the score levels. The item proved to be very difficult for most students, with only three percent receiving a score of *Complete*. Over one-half of the students in the advanced science study received a score of *Unsatisfactory*, and one-quarter omitted the question all together.

Table 3.7 ***Water at standard temperature and pressure: Percentage at each score level: 1996***

THE NATION'S REPORT CARD naep

Unsatisfactory	*Partial*	*Complete*	Omit
55	14	3	25

NOTE: Percentages may not add to 100 due to rounding.

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

Ionization Energy

This extended constructed-response question required students to apply their understanding of ionization energy to the periodic table (a copy of which was included in the chemistry block). It was scored according to a four-point scale. In order to receive a score of *Complete*, students had to choose sodium as having the lowest ionization energy and explain that a valence electron becomes more difficult to remove as one goes across period 3 because of the increasing attraction between electrons and nucleus caused by increasing effective nuclear charge. The student also had to choose francium and explain that the single valence electron becomes easier to remove as one goes down group 1A because of the decreasing attraction between nucleus and electrons due to increasing distance from the nucleus. A score of *Essential* was given to student responses that identified at least one correct element and explained why it had the lowest first ionization energy. Students who chose one or two correct elements without providing a correct explanation were given a score of *Partial*. A score of *Unsatisfactory* was given to responses that clearly demonstrated no understanding of first ionization energy.

Answer the following questions using the periodic table of the elements provided at the beginning of this section.

(a) What element in period 3 has the lowest first ionization energy? Explain why.

____

____

____

____

(b) What element in group 1A has the lowest first ionization energy? Explain why.

____

____

____

____

____

Sample 1: Ionization Energy

> Sodium. Sodium has the largest atomic radius, the lowest electronegativity, and its 3s¹ electron experiences low effective nuclear charge due to a low number of protons. Also, it needs to lose an electron to get an octet.
>
> Francium. It has the largest at rad., the lowest electronegativity, and its outermost electron experience low eff nuclear charge due to very high electron shielding.

Sample 1: *Complete* Response

The first sample response received a score of *Complete*. The student clearly understood why sodium and francium had the lowest first ionization energies in the period and group specified in the question. The student talked about the size of the atoms and the attraction between nucleus and electrons. These are part of the explanation of why sodium and francium have the lowest first ionization energies of their respective period and group.

Sample 2: Ionization Energy

> Na, because it has 1 electron in its valence shell and requires very little energy to remove that electron to make it Na^+ ion
>
> Because Fr is so large, the electron in its outermost shell is not strongly attracted to the nucleus, and therefore requires little energy to remove it.

Sample 2: *Essential* Response

The second sample response received a score of *Essential*. This student chose the correct elements, but failed to explain sodium completely. The student knew that there was only one electron in the valence shell, but did not explain why it took very little energy to remove it. The second explanation was somewhat fuller in that the student recognized that francium is large and that the electron in its outermost shell is not strongly attracted.

Sample 3: Ionization Energy

> Na it has the least mass.
>
> Fr, it mass is so large

Sample 3: *Partial* Response

The next sample shows a typical response that received a score of *Partial*. The student chose sodium and francium, but did not give a clear explanation for the choices other than that it had something to do with "mass."

Sample 4: Ionization Energy

> Ar would have the least because ionization decreases across the table.
>
> H because ionization increases as it goes down the chart + H is at the very top.

Sample 4: *Unsatisfactory* Response

The last sample response for this question received a score of *Unsatisfactory*. This student thought that ionization energy decreases across the periodic table and increases down the periodic table. In fact the reverse is true.

Table 3.8 shows the percentage of students at each of the score levels. Three percent of students received a score of *Essential* or better. Fifty-seven percent received a score of *Partial*, indicating that they could pick the correct elements, but were not able to explain why. Nearly one-third of those in the advanced science study received a score of *Unsatisfactory*.

Table 3.8 — ***Ionization energy: Percentage at each score level: 1996***

Unsatisfactory	*Partial*	*Essential*	*Complete*	Omit
30	57	3	0†	8

NOTE: Percentages may not add to 100 due to rounding.

†Percentage is between 0.0 and 0.5.

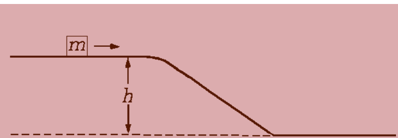
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Physics

Of the 22 physics questions included in the advanced science study, 16 were unique to the study and 6 appeared in the main NAEP assessment in science. Two multiple-choice questions (one of which was common to both the main assessment and advanced science study), one short constructed-response question, and one extended constructed-response question are described in this section.

Kinetic Energy of Block

The multiple-choice question shown below asked students to recognize the relationship between work and kinetic energy. Specifically they were asked what the kinetic energy of a block was after it had slid to the bottom of a ramp. In order to answer this question correctly, students had to know that the final kinetic energy minus the initial kinetic energy is equal to the net work (work-energy theorem). Thus the final kinetic energy, that is the energy at the bottom of the ramp, is equal to the initial energy (K) plus work (W).



A block m slides on a horizontal frictionless surface, as shown above, with a kinetic energy K. The block then slides down a frictionless ramp. The force of gravity does work W on the block as it drops a height h to a second horizontal surface. The kinetic energy of the block as it reaches the bottom of the ramp is

- (A) $2K$
- (B) $K - W$
- ● $K + W$
- (D) $W - K$

Table 3.9 contains the percentage of students choosing each response option. Sixty percent of the advanced science study students gave the correct response, and about one-quarter of the students selected option B.

Table 3.9 ***Kinetic energy of block: Percentage choosing each response: 1996***

THE NATION'S REPORT CARD

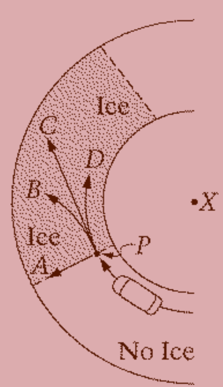
A	B	C	D	Omit
10	22	60	7	1

NOTE: Percentages may not add to 100 due to rounding.

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

Path of Car on Ice

This multiple-choice question appeared in both the advanced science study and in the main NAEP assessment, where it was classified under the physical science topic "Motion." The diagram depicts a car that is traveling around a curve onto ice. The question asks students to choose from four options which path the car is most likely to take on the ice. Students were told that the frictional force on the tires was reduced to zero. To answer this question correctly, students had to apply Newton's first law of motion, which states that an object stays at rest or continues to move at a constant speed in a straight line unless acted upon by a resultant force. In this case, the frictional force caused by the tires gripping the road is necessary for the car to move in a circle. When that force disappears, the car would continue on the path it was taking when it encountered the ice.



A car initially travels with constant speed around a tight, unbanked curve in a circular arc with center *X*, as shown in the diagram. At position *P*, the car encounters a patch of ice, which reduces the frictional force on the tires to zero.

Which of the following best shows the path that the car takes while it is on the ice?

- Ⓐ *A*
- Ⓑ *B*
- ● *C*
- Ⓓ *D*

Table 3.10 shows the percentage of students in the advanced science study choosing each response. More than half of the students gave the correct response. Option B was the most attractive of the incorrect options.

Table 3.10 ***Path of car on ice: Percentage choosing each response: 1996***

A	B	C	D	Omit
3	26	64	7	0†

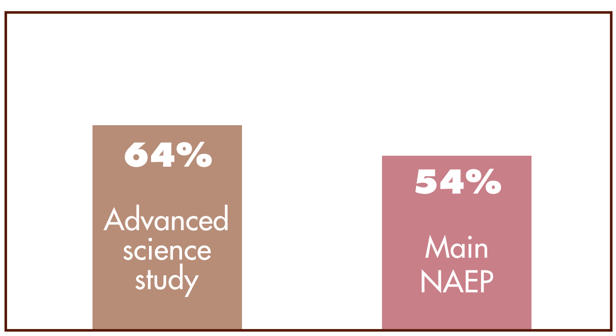
NOTE: Percentages may not add to 100 due to rounding.

†Percentage is between 0.0 and 0.5.

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

Figure 3.2 shows the percentage of students choosing the correct option in the advanced science study and the main NAEP assessment. Sixty-four percent of all students in the advanced science study chose the correct response. In main NAEP, 54 percent answered correctly.

Figure 3.2 ***Advanced science study and main NAEP— Path of car on ice: Percentage choosing the correct response: 1996***



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

Length of String

The following short constructed-response question asked students to describe how they would determine the length of a piece of string given certain equipment. It was scored using a three-point scale. A number of methods could have been used. For example, students could attach one end of the string to a weight and swing it as a pendulum to find the period, then use the equation $T=2\pi\sqrt{L/g}$ to find the length (T is the period, defined as the time it takes the pendulum to swing from the point it is released to the opposite side and back to the starting position; L is the length of the string; and g is the acceleration due to Earth's gravity, 9.8 m/s^2. A second method would involve dropping a weight through a distance equal to the length of string and measuring the time it takes to fall. The length of string could then be calculated from $L = 0.5\ at^2$ where L is the length of the string; a is the acceleration due to gravity or 9.8 m/s^2, because the weight is in free fall; and t is the time it takes the weight to fall. Students who were able to describe either of these two methods received a score of *Complete*. A score of *Partial* was given to responses that had some of the elements of a correct methodology but failed, for example, to state how the measurements could be used. A score of *Unsatisfactory* was given to responses that demonstrated no understanding of how to find the length of the string.

You are given a string about 1 meter long, but you need to determine its length more precisely. You do **not** have access to a measure tape or meter stick, but you do have access to a stopwatch, an equal-arms balance, and several different small weights with known masses. You also know that the acceleration of gravity at your location is 9.8 meters per second squared.

Describe one way in which you could determine the length of the string using the equipment available. You may not need to use all the available equipment.

Sample 1: Length of String

You could attach a weight to the string and then time the period of the object as it makes its path around the radius of the string. You could the use the equation $T = 2\pi\sqrt{L/g}$ where T = period + $g = 9.8$ then solve for l = length of string

Sample 1: *Complete* Response

The first sample response received a score of *Complete*. This student understood that the length of the string could be found using the period.

Sample 2: Length of String

Tie the weights to the end of the string. Drop the weight and hold one end of the string. Use the stopwatch to determine how long the weights took to drop. Then square the time, multiply by 1/2, and ~~and~~ then multiply a 9.8 ($\Delta D = \frac{1}{2} A(T^2)$)

Sample 2: *Complete* Response

The student responsible for the next response (also scored *Complete*) used a different method to find the length of the string. Here the student tied the weight to the end of the string and timed how long it took to fall. The student then used the formula $L=(1/2)at^2$ to determine the length of the string, showing the correct use of the acceleration due to gravity, 9.8 m/s^2.

Sample 3: Length of String

You can use the weight to create a pendulum and use the stopwatch to find the period of the pendulum + then calculate the length of the string.

Sample 3: *Partial* Response

The third sample response received a score of *Partial*. The student clearly states that a pendulum can be created, but fails to state how the measurements can be used to find the length of the string, merely stating "then calculate the length of the string."

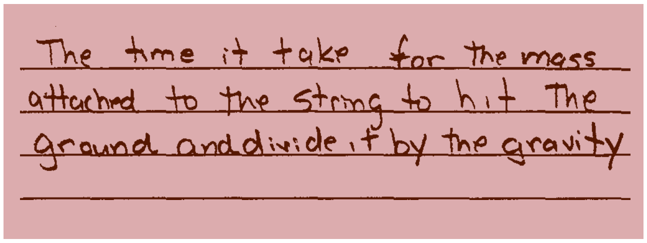
Sample 4: Length of String

The length of the string can be determined by timing how long it takes a small weight of known mass to fall the distance of the string.

Sample 4: *Partial* Response

The fourth sample response also received a *Partial* rating. The student chooses to time how long it takes a weight to fall the distance of the string but, again, does not tell how to use the data.

Sample 5: Length of String


The time it take for the mass attached to the string to hit The ground and divide it by the gravity

Sample 5: *Unsatisfactory* Response

The next sample shows a typical response that received a score of *Unsatisfactory*. This student has some elements of a method, but the explanation is unclear.

Table 3.11 presents the percentage of students at each of the score levels. The item proved to be a challenge for most students, with 50 percent of the students in the advanced science study providing an *Unsatisfactory* response and 22 percent of the students omitting the question altogether.

Table 3.11 ***Length of string: Percentage at each score level: 1996***

THE NATION'S REPORT CARD naep

Unsatisfactory	*Partial*	*Complete*	Omit
50	20	4	22

NOTE: Percentages may not add to 100 due to rounding.

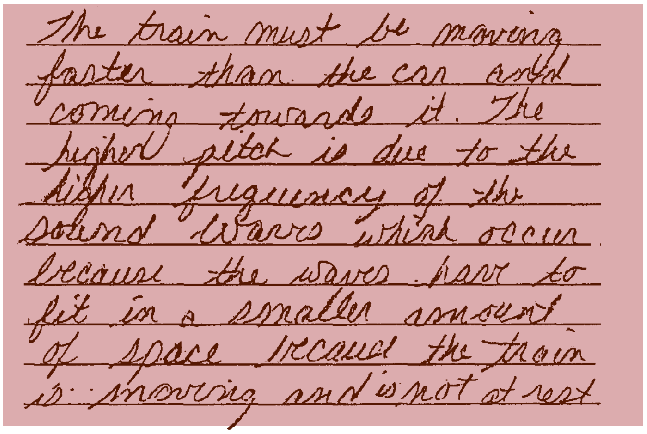
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Direction and Speed of Train Relative to Car

The next sample is an extended constructed-response question that asked students to infer from a given observation the direction of motion and speed of a train and to support their inference with an explanation. The question was scored using a four-point scale. In order to receive a score of *Complete*, students needed to (1) state that the train was moving toward the car, (2) state that the speed of the train was greater than that of the car, and (3) provide an explanation correctly related to the Doppler effect, a decrease in wavelength, an increase in frequency, or a compression of the waves arriving at the car. A score of *Essential* was given to those responses that met two of the criteria, and a score of *Partial* was given to those responses that met one criterion only. Responses were given a score of *Unsatisfactory* when they met none of the criteria.

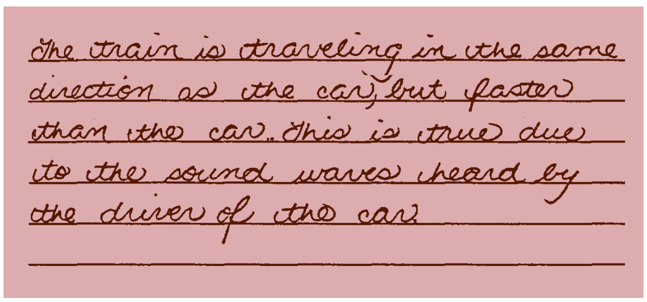
A car travels at a constant speed along a road that is adjacent and parallel to a railroad track. A train is behind the car. The driver of the car hears the sound of the train whistle, but the pitch of the whistle is higher than would have been heard if the train and the driver were at rest. What, if anything, can be inferred from this observation alone about the direction of motion of the train and the speed of the train? Explain your reasoning.

Sample 1: Speed of Train


The train must be moving faster than the car and coming towards it. The higher pitch is due to the higher frequency of the sound waves which occur because the waves have to fit in a smaller amount of space because the train is moving and is not at rest

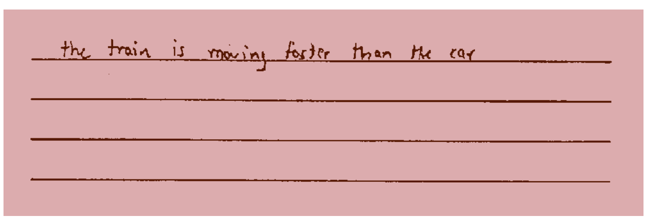
Sample 1: *Complete* Response
The first sample response received a score of *Complete*. This student understood that the train was moving faster than the car and toward it, and also gives an explanation based on the Doppler effect. Specifically, the student recognizes that the higher pitch is due to the higher frequency of sound waves reaching the car, which results from the shortening of their wavelength due to movement of the train.

Sample 2: Speed of Train


The train is traveling in the same direction as the car, but faster than the car. This is true due to the sound waves heard by the driver of the car

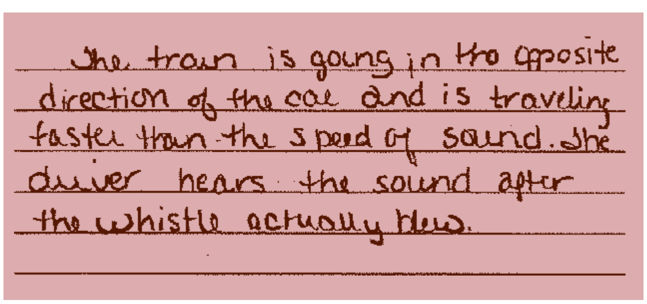
Sample 2: *Essential* Response
The second sample received a score of *Essential*. The student was able to say that the train was “traveling in the same direction as the car, but faster than the car,” but was not able to explain why, other than to state “This is true due to the sound waves heard by the driver of the car.”

Sample 3: Speed of Train


the train is moving faster than the car

Sample 3: *Partial* Response
The third sample response received a score of *Partial*. The student knew that “the train is moving faster than the car,” but failed to elaborate further.

Sample 4: Speed of Train


The train is going in the opposite direction of the car and is traveling faster than the speed of sound. The driver hears the sound after the whistle actually blew.

Sample 4: *Unsatisfactory* Response
The fourth sample shows a typical response that received a score of *Unsatisfactory*. This student believed that the train was moving away from the car and was traveling faster than the speed of sound.

Table 3.12 presents the percentage of students at each of the score levels. Fifty-two percent of the advanced student population received an *Unsatisfactory* score. The top score—*Complete*—was achieved by less than ten percent of students. Thirty-nine percent of the students received either a *Partial* or *Essential* score.

Table 3.12 **Direction and speed of train relative to car: Percentage at each score level: 1996**

THE NATION'S REPORT CARD naep

Unsatisfactory	*Partial*	*Essential*	*Complete*	Omit
52	20	19	7	0†

NOTE: Percentages may not add to 100 due to rounding.

†Percentage is between 0.0 and 0.5.

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

Item Maps of Sample Questions

Up to this point, the results related to students' performance on individual questions have been reported as the percentage of students in each of the possible response categories. Figure 3.3 provides a visual representation of the relative difficulty of each question in relation to other questions by mapping them on the 0 to 300 scale. The left side of the figure maps the chemistry and physics sample questions, and the right side maps the biology sample questions. Each multiple-choice question is denoted by (mc), and the short descriptions for the constructed-response questions are each followed by the score value in parentheses: 3 (*Complete*) and 2 (*Partial*) for questions with a three-level scoring guide, or 4 (*Complete*), 3 (*Essential*) and 2 (*Partial*) for a four-level scoring guide. 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 four-option, multiple-choice question.[1] Therefore, the further down the question appears on the scale, the easier it was, and the higher it maps on the scale, the more difficult it was.

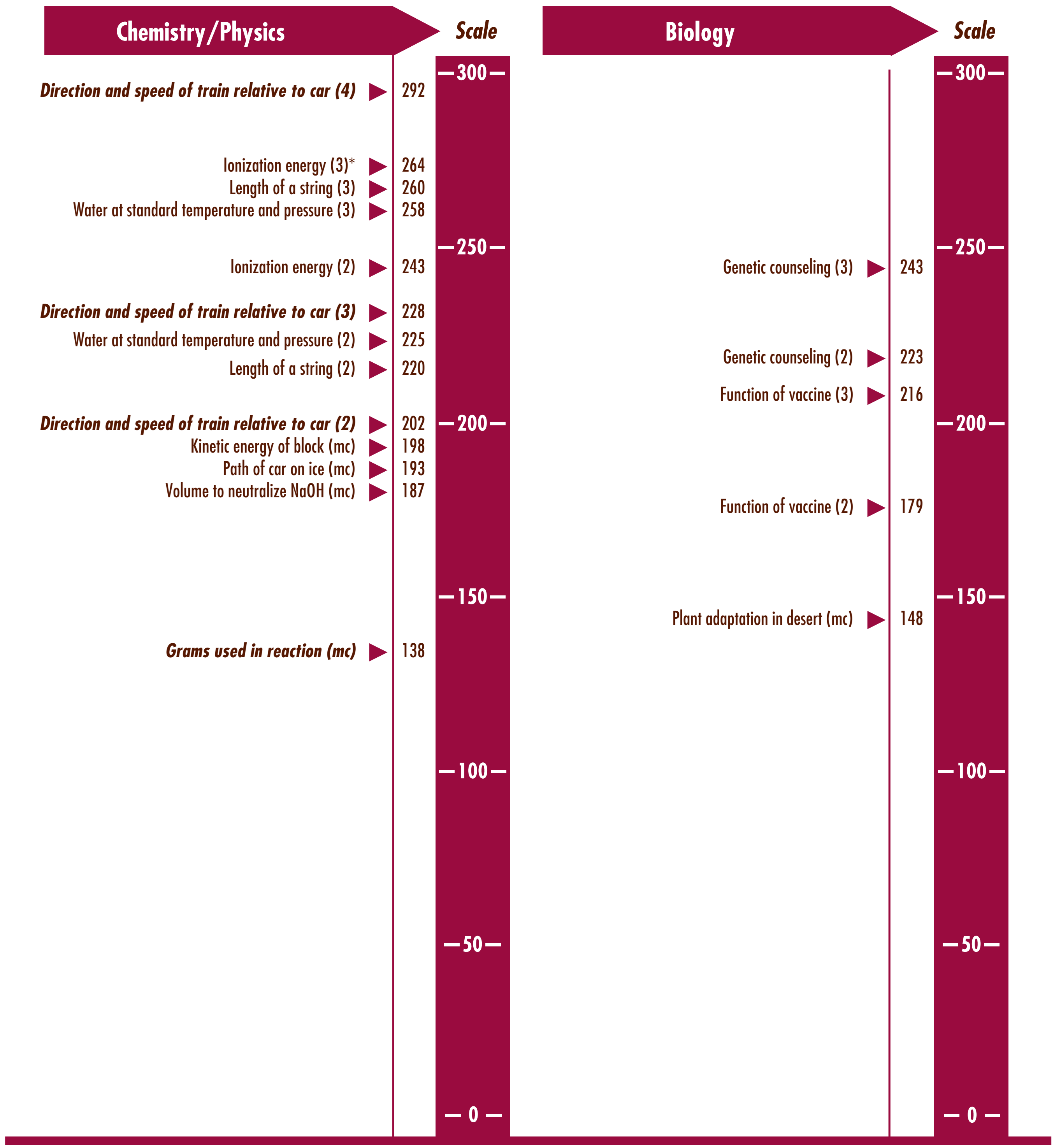
The following examples will help in interpreting these maps. The multiple-choice question about number of grams used in a reaction maps at the 138 point on the chemistry/physics scale. This means that twelfth-grade students in the advanced science study with science scale scores at or above 138 have at least a 74 percent chance of answering this question correctly. Put slightly differently, this question is answered correctly by at least 74 of every 100 students scoring at or above the 138 scale score level. This does not mean that students at or above the 138 scale score always answer the question correctly or that students below the 138 scale score always answer it incorrectly. Rather, students have a higher or lower probability of successfully answering the question depending on their overall ability as measured on the advanced science study scale. As another example, consider the constructed-response question that maps at a scale score of 292 on the chemistry/physics scale. Scoring of responses to this question allows for partial credit by using a four-level scoring guide. Mapping the question at the 292 scale score indicates that at least 65 percent of the students performing at or above this point achieved a score of 4 (*Complete*) on the question. Students whose overall performance falls close to 228 have a 65 percent chance of receiving a score of 3 (*Essential*) on the same question, and those who fall around a scale score of 202 have a 65 percent chance of receiving a score of 2 (*Partial*).

---

[1] The use of a higher criterion for multiple-choice questions reflected students' ability to guess the correct answer from among the alternatives.

Figure 3.3

Map of selected sample questions on the NAEP science scale for advanced science study: 1996



NOTE: Position of questions is approximate. The average scale scores were 175 for chemistry/physics and 173 for biology.

Each multiple-choice question is denoted by (mc), and the short descriptions for the constructed-response questions are each followed by the score value in parentheses: 3 (*Complete*) and 2 (*Partial*) for questions with a three-level scoring guide, or 4 (*Complete*), 3 (*Essential*) and 2 (*Partial*) for a four-level scoring guide.

*The sample size was insufficient to estimate a scale score for a complete response on the ionization question.

Performance on Common Questions

Each student who participated in the advanced science study was presented with 18 questions from the 1996 main NAEP science assessment. While student performance on the advanced science assessment has been analyzed in a number of different ways, comparisons with results obtained from the main assessment can only be made at the item level on the set of questions common to both the study and the main assessment. Since most of the common items will be used in the NAEP 2004 science assessment, they cannot be discussed here in any detail. Table 3.13 shows the question score obtained by students who participated in the advanced science study and students from the main assessment for questions that were common to both. For multiple-choice questions, the question score is identical to the percentage of students answering the question correctly. For constructed-response questions that were scored on either a three- or four-level scale, the question score represents the average of all student scores on that question expressed as a percentage of the maximum possible score. Missing responses

Table 3.13 ***Percentage correct on common items for students in the advanced science study and main NAEP: 1996***

THE NATION'S REPORT CARD naep

Content Domain	Item type[1]	Question score in advanced study*	Question score in main NAEP**
Biology			
Theory of Evolution	mc	.71	.53
Major Plant Group	mc	.45	.35
Evolutionary Relationships	mc	.33	.31
Temperature Regulation	scr	.57	.49
Cause of Menstruation	scr	.20	.13
Research Project	ecr	.35	.26
Chemistry			
Stoichiometry	mc	.88	.76
Exothermic Reaction	mc	.69	.57
Ionic Properties	mc	.57	.41
Neutralization	scr	.41	.22
Test for pH	scr	.41	.28
Rate of Movement	scr	.32	.25
Physics			
Acceleration	mc	.89	.74
Nuclear Decay	mc	.73	.59
Path of Car on Ice	mc	.64	.54
Electrical Circuits	scr	.57	.47
Predict Composition of Object	scr	.25	.22
Devise Density Experiment	ecr	.37	.23

[1] mc = multiple-choice, scr = short constructed-response, ecr = extended constructed-response

* Question score obtained by students who participated in the advanced science study.

**Question score obtained by all students who took part in the main NAEP science assessment.

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

before the last observed response in a block were considered intentional omissions and were included in the denominator of the statistic. Students in the advanced science study were more likely than the students in the national sample to respond correctly to the set of common questions administered to both groups.

Summary of Student Performance on Sample Questions

The data presented in this chapter indicate how students performed on selected questions taken from the advanced science study. Discussion of individual questions was limited to those that have been released to the public. Two questions that appeared in both the advanced science study and the main NAEP assessment have been released to the public and were discussed in this chapter.[2]

- Students in the advanced science study were more likely than the students in the national sample to respond correctly to the set of common questions administered to both groups. The difference in question scores between the advanced study and main NAEP samples on common questions ranged from 2 percent to 19 percent.
- In general, constructed-response questions were more difficult than multiple-choice questions, and constructed-response questions tended to have a higher percentage of omits than multiple-choice questions. This was also true for the main NAEP assessment.

---

[2] National Center for Education Statistics, National Assessment of Educational Progress. (1997). *1996 science assessment public release, grade 12.* [on-line]. Available: http://nces.ed.gov/nationsreportcard.

Appendix A

Scoring Guides for Sample Questions from the Advanced Science Study

Biology Content Domain

Function of vaccine question

Children must be immunized against childhood diseases such as measles and mumps in order to attend public school. Explain how the vaccine used for one of these immunizations confers immunity.

____

____

____

____

____

Scoring guide

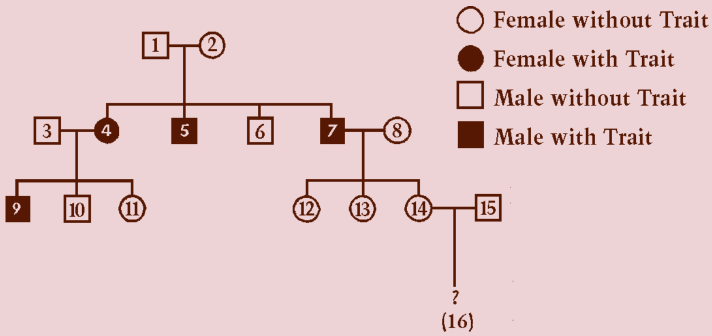
(3) Complete	Student response clearly indicates how immunizations confer immunity.
(2) Partial	Student response provides a partial description of immunity.
(1) Unsatisfactory/Incorrect	Student response fails to demonstrate a scientific understanding of immunity.

Guidelines for Crediting Response

Vaccines contain pieces of infectious agent (or heat-killed infectious agents). Vaccines cause the body to produce antibodies. When an infection takes place, the immune system recognizes the infectious agent and is able to rapidly produce antibodies against it. These antibodies destroy the virus before the infection can proceed.

Genetic counseling question

A young woman who is pregnant seeks genetic counseling. She wants to know whether her child is likely to be born with a genetic trait that is common in her family, though it has not been observed in her husband's family. The counselor records the occurrence of the trait in her family and constructs the following pedigree. The woman is represented by 14 in the pedigree.



Will the woman's child (16 in the pedigree) have the trait? Explain the reasoning behind your conclusion.

____

____

____

____

____

____

Scoring guide

(3) Complete	Student response concludes that the child may or may not show the trait and correctly explains the reasoning for either conclusion.
(2) Partial	Student response concludes that the child is unlikely to show the trait, but does not clearly explain the evidence for this.
(1) Unsatisfactory/Incorrect	Student response indicates no understanding of the concepts.

Guidelines for Crediting Response

The child will probably not show the trait since the father is unlikely to carry the recessive allele for the trait; therefore, even if the young woman carries the trait the child will not express it. Alternatively, the child may express the recessive allele but only if both the father and the mother are carriers of the recessive allele. It is however unlikely that the father carries the recessive allele since there is no evidence of the trait in his family.

Chemistry Content Domain

Water at standard temperature and pressure question

Because of its low molecular mass, H_2O would be expected to be a gas at room temperature and normal atmospheric pressure. Explain in terms of chemical bonding and intermolecular forces why it is a liquid under these conditions and not a gas.

Scoring guide

(3) Complete	Student response correctly explains why water is a liquid at room temperature through discussion of chemical bonding and intermolecular forces.
(2) Partial	Student response correctly explains why water is a liquid at room temperature through discussion of chemical bonding **or** intermolecular forces.
(1) Unsatisfactory/Incorrect	Student response indicates little or no understanding of the reason why water is a liquid at room temperature.

Guidelines for Crediting Responses

H_2O molecules are polar, (or dipoles) thus the intermolecular forces among H_2O molecules are strong because of the dipole-dipole interactions present among H_2O molecules; or hydrogen bonding is a strong type of intermolecular force and these strong attractions hold the molecules together tightly at room temperature.

Ionization of energy question

Answer the following questions using the periodic table of the elements provided at the beginning of this section.

(a) What element in period 3 has the lowest first ionization energy? Explain why.

____

____

____

____

____

(b) What element in group 1A has the lowest first ionization energy? Explain why.

____

____

____

____

____

Scoring guide

(4) Complete	Student response demonstrates knowledge of trends in the periodic table by identifying elements that have the lowest ionization energy and explaining why their ionization energy is the lowest.
(3) Essential	Student response identifies one or two elements that have the lowest ionization energy and explains why the ionization energy of one of the elements is the lowest.
(2) Partial	Student response identifies one or two elements that have the lowest ionization energy, but is unable to explain why.
(1) Unsatisfactory/Incorrect	Student response demonstrates no knowledge of trends in the periodic table.

Guidelines for Crediting Responses

a) Na (sodium) because, in going across period 3, a valence electron becomes more difficult to remove because of increasing attraction due to the increasing effective nuclear charge.

b) Fr (francium) because, in going down group IA, the single valence electron becomes easier to remove because of the decreasing attraction due to increasing distance from the nucleus.

Physics Content Domain

Length of string question

You are given a string about 1 meter long, but you need to determine its length more precisely. You do **not** have access to a measuring tape or meter stick, but you do have access to a stopwatch, an equal-arms balance, and several different small weights with known masses. You also know that the acceleration of gravity at your location is 9.8 meters per second squared.

Describe one way in which you could determine the length of the string using the equipment available. You may not need to use all the available equipment.

____

____

____

____

____

Scoring guide

(3) Complete	Student response describes either the pendulum method or the time of fall method.
(2) Partial	Student response demonstrates partial understanding of either the pendulum method or the time of fall method.
(1) Unsatisfactory/Incorrect	Student response does not describe any correct method.

Guidelines for Crediting Responses

Method 1:
Attach a weight to one end of the string of length L, swing it as a pendulum, and measure the period (T). The equation is: $T=2\pi\sqrt{L/g}$, where g = 9.8 m/sec^2. Solve for L.

Method 2:
Drop a weight through a distance equal to the length of the string (L), and measure the time of fall (t). The equation is: $L=(1/2)\ at^2$, where a is the acceleration due to gravity or 9.8 m/s^2.

Note: The pendulum method is experimentally better, but full credit should be given for method 2 because it is theoretically correct and could be done experimentally with precision with more accurate timing equipment.

Direction and speed of train relative to car question

A car travels at a constant speed along a road that is adjacent and parallel to a railroad track. A train is behind the car. The driver of the car hears the sound of the train whistle, but the pitch of the whistle is higher than would have been heard if the train and the driver were at rest. What, if anything, can be inferred from this observation alone about the direction of motion of the train and the speed of the train? Explain your reasoning.

____

____

____

____

____

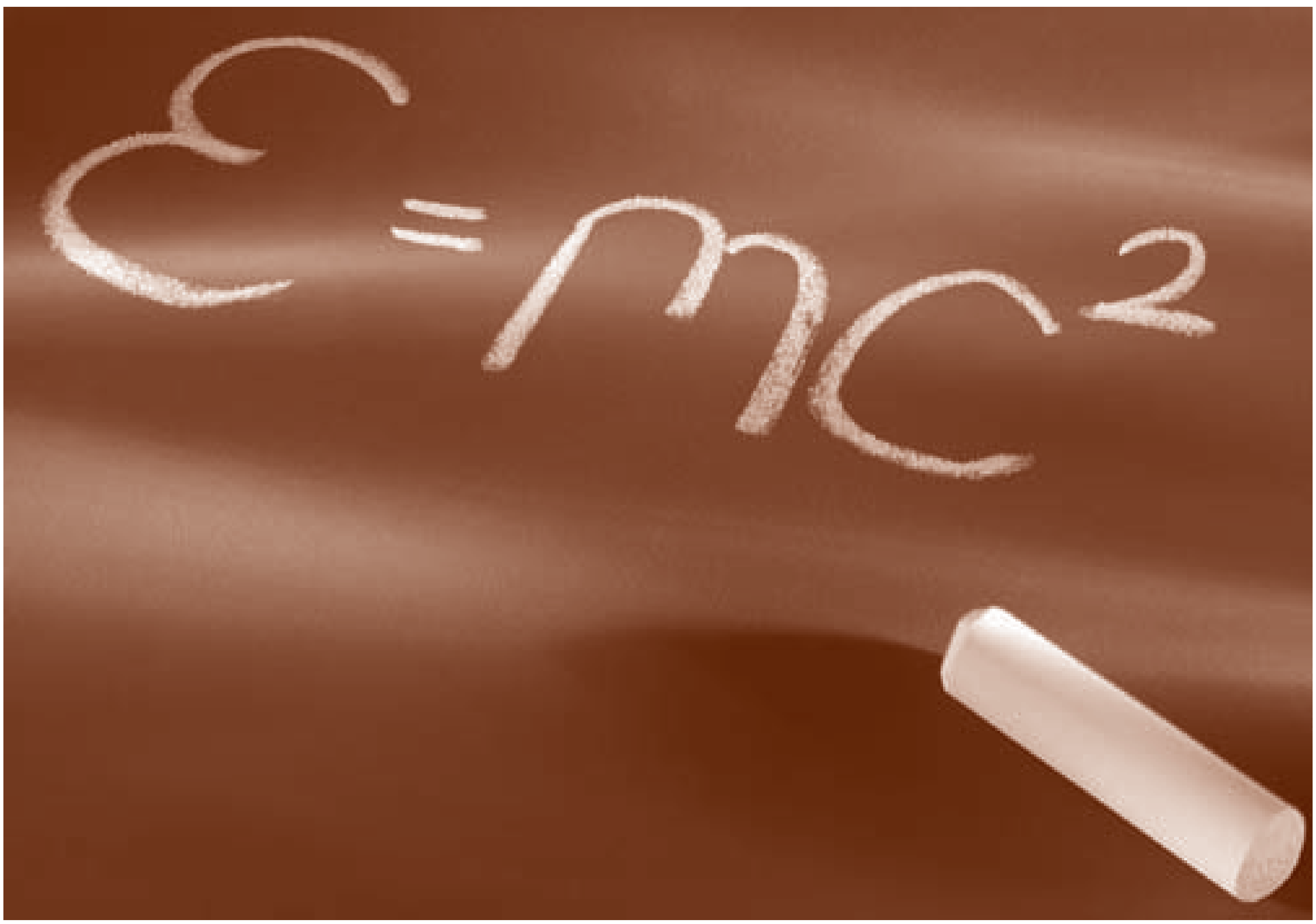
Scoring guide

(4) Complete	Student response (1) indicates that the train is moving toward the car or getting closer to the car, (2) indicates that the speed of the train is greater than that of the car, and (3) gives a reasonable explanation.
(3) Essential	Student response makes any 2 of the 3 statements above required for a complete answer.
(2) Partial	Student response indicates that the train is moving toward the car but gives no explanation. **–or–** Student response indicates that the speed of the train is greater than that of the car but gives no explanation. **–or–** Student response has some correct description relating to the Doppler effect, but makes no correct inference about the direction of motion or speed of the train.
(1) Unsatisfactory/Incorrect	Student response fails to indicate either that the train is moving toward the car or that the speed of the train is greater than the car, and fails to provide any correct description relating to the Doppler effect.

Guidelines for Crediting Responses

There are three elements that comprise a complete response:

1. The train is moving in the direction of the car (or train is moving toward the car).
2. The speed of the train was greater than that of the car.
3. A correct explanation relating to the Doppler effect, a decrease in wavelength, an increase in frequency, or a compression of the waves arriving at the car.


E=mc2

Appendix B

Overview of Procedures Used for the NAEP Advanced Science Study

The NAEP 1996 Science Assessment

A number of reports have been written addressing different aspects of the NAEP 1996 science assessment. The *NAEP 1996 Science Report Card for the Nation and the States*, for example, presents the results related to what various subgroups of students in grades 4, 8, and 12 know and can do in science.[1] Another report designed to meet the information needs of teachers focuses on sample questions and student responses.[2] A report written for policymakers and administrators includes results related to teacher preparation, the amount of emphasis science instruction receives in schools, and the availability of school resources that support science instruction.[3] The information provided in these and other related reports can be useful in putting the results of the advanced science study into a broader context. Highlights from the various science reports can be found on the World Wide Web at http://nces.ed.gov/nationsreportcard/science

The advanced science study differed from the main NAEP assessment with respect to certain characteristics of the students selected to participate and how the questions were categorized. While the students who took the main NAEP assessment represented the general population of fourth-, eighth-, and twelfth-graders, only twelfth-grade students who were enrolled in an advanced science course were eligible to participate in the advanced science study. The questions in both assessments were based on the *Science Framework for the 1996 and 2000 National Assessment of Educational Progress*;[4] however they were categorized as either life, physical, or earth sciences in the main NAEP assessment and biology, chemistry, or physics in the advanced science study. Forty-eight of the 66 questions in the advanced science

---

[1] O'Sullivan, C.Y., Reese, C.M., & Mazzeo, J. (1997). *NAEP 1996 science report card for the nations and the states*. Washington, DC: National Center for Education Statistics.

[2] O'Sullivan, C.Y., & Weiss, A.R. (1999). *Student work and teacher practices in science*. Washington, DC: National Center for Education Statistics.

[3] O'Sullivan, C.Y. Weiss, A.R., & Askew, J.M. (1998). *Students learning science: A report on policies and practices in U.S. schools*. Washington, DC: National Center for Education Statistics.

[4] National Assessment Governing Board. (1995). *Science framework for the 1996 and 2000 National Assessment of Educational Progress*. Washington, DC: Author

study were unique to that assessment and were developed to assess more challenging material than the main NAEP assessment. The other 18 questions were taken from the main NAEP assessment to allow for some comparisons in student performance. Students participating in the main NAEP science assessment and in the advanced science study were assessed between January 3 and March 29, 1996.

The purpose of this section is to provide a general overview of the procedures used for the advanced science study. A more detailed description of the procedures for the NAEP 1996 science assessment can be found in the *NAEP 1996 Technical Report*.[5]

The Advanced Science Study

Students who participated in the advanced science study each received one of three assessment booklets containing seven sections, or blocks. Four of the blocks contained cognitive questions and three blocks contained background questions including general background questions, science background questions, and questions about students' motivation to do well on the assessment.

The cognitive block structure of the advanced science study booklets is presented in table B.1. Each booklet started with a linking block (S19) containing the 18 questions taken from various blocks used in the main assessment. This was followed by three advanced blocks: 16 questions in biology (S16); 16 questions in chemistry (S17); and 16 questions in physics (S18). Each of the advanced blocks contained 7 multiple-choice and 9 constructed-response questions. The question order within each block remained the same regardless of the order of the blocks in the booklet. Students were given 30 minutes to complete each of the four cognitive blocks, for a total of two hours compared to the one-and-a-half hours allotted for the main NAEP assessment, which contained three cognitive blocks.

Table B.1 ***Block structure of the advanced science study booklets***

THE NATION'S REPORT CARD naep

Booklet	Cognitive blocks
S238	S19, S18, S17, S16
S239	S19, S17, S16, S18
S240	S19, S16, S18, S17

[5] Allen, N.L., Carlson, J.E., & Zelenak, C.A. (1999). The *NAEP 1996 technical report*. Washington, DC: National Center for Education Statistics.

Sampling Procedures for the Advanced Science Study

The population for the advanced science study was sampled as part of the complex multistage sample design used for the main national sample. The design involved sampling students from selected schools within selected geographic areas across the United States. The sample design had the following stages:

1. Selection of primary sampling units (geographic areas such as a county, a group of counties, or metropolitan statistical area);
2. Selection of schools (public and nonpublic) within selected areas; and
3. Selection of students within the selected schools.

Table B.2 shows the number of students per school who were assessed in the main NAEP and advanced science study. In addition to representing the respective populations as a whole, the main assessment and advanced science study samples involved oversampling of nonpublic schools, and of public schools with a moderate or high enrollment of black or Hispanic students. This oversampling was undertaken to increase the sample sizes of nonpublic school students and minority students, so as to increase the reliability of estimates for those groups of students.

Table B.2 ***Number of students per school for the main NAEP assessment and advanced science study***

THE NATION'S REPORT CARD NAEP

Sample	Number of assessed students	Number of schools	Mean number of students per assessment per school	Mean number of students per item per school
Grade 4				
Main sample	11,578	421	27.5	4.5–7.4*
Grade 8				
Main sample	11,971	346	34.6	5.6–9.4*
Grade 12				
Main sample	11,481	401	28.6	4.6–7.7*
Advanced science sample	2,431	222	11.0	11.0

* This number varied because some item blocks appeared more often than others in the set of booklets used for this sample.

As part of the ongoing effort to report results that represent the performance of all students, the 1996 main NAEP assessment was conducted using revised criteria for the inclusion of students with disabilities and limited English proficient (SD/LEP) students. A detailed description of the process used to include these students can be found in chapter 4 of the *NAEP 1996 Science Report Card for the Nation and the States*[6]. Like the main NAEP, the results of the advanced science study were reported for a sample of students selected using criteria that emphasized inclusion as opposed to exclusion. Although the results of the 1996 main NAEP assessment included some supplemental data for students who received some type of assessment accommodation (e.g., one-on-one testing, extended time, oral reading, or signing of directions), the information provided in this report is for non-accommodated students only.

Participation Rates

In order for the data to be reported, NCES school and student participation rates must be met. School nonparticipation or student nonparticipation in the form of absenteeism creates a potential for bias to be introduced in the reporting of the data. The participation rates of schools and students included in the 1996 assessments were inspected for any differences. NCES standards regarding acceptable potentials for bias are expressed in terms of weighted participation rates. Table B.3 shows the weighted participation rates by grade and session type for the main NAEP and advanced science samples. The overall participation rates reflect the extent to which both the desired school and student participation were achieved.

Table B.3 ***Weighted participation rates (in percentages), by grade for 1996 main NAEP and advanced science reporting samples***

THE NATION'S REPORT CARD naep

	Main NAEP science	Advanced science
Grade 4		
School participation	77.8	—
Student participation	94.9	—
Overall participation	73.8	—
Grade 8		
School participation	79.7	—
Student participation	93.1	—
Overall participation	74.3	—
Grade 12		
School participation	77.4	77.7
Student participation	77.5	86.5
Overall participation	60.0	67.2

[6] O'Sullivan, C.Y., Reese, C.M., & Mazzeo, J. (1997). *NAEP 1996 science report card for the nation and the states*. Washington, DC: National Center for Education Statistics.

Evaluation of Potential for Bias

Although school and student nonresponse adjustments are intended to reduce the potential for nonparticipation to bias the assessment results, they cannot completely eliminate this potential bias with certainty. The extent of bias remains unknown because there are no assessment data for the nonparticipating schools and students.

Some insight can be gained about the potential for residual nonresponse bias by examining the weighted school- and student-level distributions of characteristics known for both participants and nonparticipants, especially for those characteristics known or thought likely to be related to achievement on the assessment. If the distributions for the full sample of schools (or students) without the use of nonresponse adjustments are close to those for the participants with nonresponse adjustments applied, there is reason to be confident that the bias from nonparticipation is small.

A nonresponse bias analysis completed on the reporting population for the science assessment can be found in the *NAEP 1996 Technical Report*.[7] Science was chosen because it contained the largest number of students and could, therefore, provide the most precise estimates of student distributions across several demographic characteristics (i.e., age, race, gender, type of school, and school size). Generally, the findings show that the student distributions before and after school and/or student nonresponse adjustments were similar, with a few exceptions. Most of these exceptions were at grade 12 due to its relatively high nonresponse rate (20.3 percent for grade 12 students).

Within the NAEP data, there are several school-level characteristics available for both participating and nonparticipating schools. The tables that follow show the combined impact of nonresponse and of the nonresponse adjustments on the distributions of schools (weighted by the estimated number of eligible students enrolled) and students, by the type of school (public, Catholic, other nonpublic), the size of the school (measured by the estimated number of eligible students enrolled), and whether the school is located in an urban or rural place.

Several student-level characteristics are available for both absent and assessed students. The tables that follow show the impact of school nonresponse and nonresponse adjustments, as well as student nonresponse and nonresponse adjustments, on the distributions of eligible students at grade 12. The distributions are presented by age category (at or below modal age, and above modal age), race/ethnicity category (white, black, Hispanic, and other), gender, and SD/LEP status.

---

[7] Allen, N.L., Carlson, J.E., & Zelenak, C.A. (1999). The *NAEP 1996 technical report*. Washington, DC: National Center for Education Statistics.

Table B.4 shows the weighted marginal distributions of students for each of the three classification variables for grade 12, using weighted eligible schools. The distributions before school nonresponse adjustments are based on the full sample of in-scope schools for each assessment—those participating, plus those refusals for which no substitute participated. The distributions after school nonresponse adjustments are based only on participating schools for each assessment, with school nonresponse adjustments applied to them. The weighted school-level nonparticipation rates at grade 12 are 22.6 percent for science and 22.3 percent for advanced science.

It can be seen from table B.4 that, overall, the distributions for school type, school size, and school location remain similar.

Table B.4 **Distribution (in percentage) of populations of eligible students based on full weighted sample of eligible schools before and after school nonresponse adjustments, 1996 main NAEP samples, grade 12**

THE NATION'S REPORT CARD naep

Population	Science		Advanced science	
	Before	After	Before	After
School Type				
Catholic	5.5	5.9	5.4	5.5
Other nonpublic	4.2	3.4	3.8	3.1
Public	90.3	90.7	90.7	91.4
School Size				
1 (1–49)	5.0	4.7	5.5	5.0
2 (50–399)	70.5	67.0	69.4	65.7
3 (400+)	24.5	28.3	25.1	29.2
School Location				
Large city	15.4	16.8	15.4	16.0
Midsize city	16.3	18.3	16.3	18.4
Urban fringe/large city	23.3	21.2	23.8	22.5
Urban fringe/midsize city	15.3	14.4	15.0	13.9
Large town	1.5	0.9	1.5	0.9
Small town	14.4	15.5	14.7	15.9
Rural MSA[1]	3.7	3.7	4.1	4.1
Rural nonMSA	10.2	9.5	9.1	8.4

[1] Metropolitan Statistical Area

Table B.5 shows the distributions of two school-level characteristics—school type and school location—plus additional distributions of student-level characteristics, using weighted eligible students. The distributions before student nonresponse adjustments are based on assessed and absent students (with base weights adjusted for school nonparticipation). The distributions after student nonresponse adjustments are based on assessed students only, with the student nonresponse adjustments also applied to them. The rates of student nonparticipation at grade 12 are 22.5 percent for science and 13.5 percent for advanced science.

Table B.5 shows that, the effect of the student nonparticipation adjustment has resulted in very little change in distribution.

Table B.5 *Distribution (in percentage) of populations of eligible students before and after student nonresponse adjustments, 1996 main NAEP samples, grade 12*

Population	Science		Advanced science	
	Before	After	Before	After
School Type				
Catholic	6.4	8.1	10.6	11.7
Other nonpublic	3.6	4.2	4.9	5.4
Public	90.0	87.6	84.5	82.9
School Location				
Large city	17.0	16.9	15.5	15.2
Midsize city	18.2	16.9	16.7	15.6
Urban fringe/large city	21.4	21.7	27.1	27.6
Urban fringe/midsize city	13.6	14.8	12.1	12.9
Large town	1.0	1.0	0.8	0.7
Small town	15.5	15.2	13.9	13.5
Rural MSA	3.9	4.1	5.6	5.9
Rural nonMSA	9.4	9.6	8.3	8.7
Age Category				
At modal age or younger	64.3	64.5	72.5	72.3
Older than modal age	35.7	35.5	27.5	27.7
Race/Ethnicity				
White	69.4	69.1	74.8	74.1
Black	12.6	12.2	8.9	9.1
Hispanic	11.0	11.1	6.8	7.0
Other	7.1	7.5	9.5	9.7
Gender[1]				
Male	48.7	48.4	49.4	49.2
Female	51.2	51.6	50.5	50.8
SD				
Yes	3.4	3.0	0.4	0.3
No	96.6	97.0	99.6	99.7
LEP				
Yes	2.1	2.2	1.1	1.0
No	97.9	97.8	98.9	99.0
SD, LEP				
SD yes; LEP yes	0.1	0.0	0.0	0.0
SD yes; LEP no	3.3	3.0	0.4	0.3
SD no; LEP yes	2.1	2.1	1.1	1.0
SD no; LEP no	94.6	94.8	98.5	98.6

[1] Gender is unknown for a small percentage of students.

Table B.6 shows the weighted distributions of eligible students in participating schools, using the base weights of assessed and absent students unadjusted for school-level nonresponse. Tables B.5 and B.6 show that both school and student-level nonresponse and nonresponse adjustments have little effect on the distributions of eligible students by age,

Table B.6 Distribution (in percentage) of populations of eligible students before school and student nonresponse adjustment, 1996 main NAEP samples, grade 12

THE NATION'S REPORT CARD naep

Population	Science	Advanced science
Age Category		
At modal age or younger	63.9	72.1
Older than modal age	36.1	27.9
Race/Ethnicity		
White	70.3	76.0
Black	12.4	8.8
Hispanic	10.3	6.3
Other	6.9	9.0
Gender[1]		
Male	48.8	49.6
Female	51.2	50.3
SD		
Yes	3.4	0.4
No	96.6	99.6
LEP		
Yes	2.0	1.0
No	98.0	99.0
SD, LEP		
SD yes; LEP yes	0.1	0.0
SD yes; LEP no	3.3	0.4
SD no; LEP yes	1.9	1.0
SD no; LEP no	94.7	98.6

[1] Gender is unknown for a small percentage of students.

race/ethnicity, gender, and SD/LEP status. All of the distributions in the tables are similar. When comparing the distributions in table B.5 before and after student nonresponse adjustments, one expects the distributions by age category and race/ethnicity to be similar because these variables were used to determine student nonresponse adjustment classes. However, the distributions by gender and SD/LEP status are also similar. To the extent that nonrespondents would perform like respondents with the same characteristics (defined by the characteristics in the tables), the bias in the assessment data is small.

Further information about potential nonresponse bias can be gained by studying the absent students. NAEP proficiency estimates are biased to the extent that assessed and absent students within the same weighting class differ in their distribution of proficiency. It seems likely that the assumption that absent students are similar in proficiency to assessed students is reasonable for some absent students—namely, those whose absence can be characterized as random. Conversely, it seems likely that students with longer and more consistent patterns of absenteeism—such as truants, dropouts, near dropouts, and the chronically ill—are unlikely to be as proficient as their assessed counterparts.

In the 1996 assessments, schools were asked to classify each absent student into one of nine categories. The results of this classification for the assessments are shown in table B.7. As anticipated, the majority of absenteeism from the assessment was the result of an absence from school of a temporary and unscheduled nature. The table shows that in the advanced session, the overall absentee rate is lower than in the 'non-advanced' session. The proportion of absenteeism classified as temporary is similar in both sessions.

In grade 12, chronic truants, those suspended, and those in school but not attending constitute the obvious candidates for potential bias. These groups comprise 6.0 percent of absent students in the advanced session, thus their potential for introducing significant bias under the current procedures is minor.

Table B.7 ***Weighted distribution (in percentage) of absent students, by nature of absenteeism, 1996 assessments grade 12***

THE NATION'S REPORT CARD

Population	Science	Advanced science
Temporary absence[1]	63.6	64.5
Long-term absence[2]	0.8	0.0
Chronic truant	1.4	0.0
Suspended or expelled	0.4	0.3
Parent refusal	9.2	11.7
Student refusal	12.0	12.9
In school, did not attend session	7.0	5.7
In school, not invited[3]	0.0	0.0
Other	3.9	4.8
Missing	1.8	0.0
Total absentee	2,269	379
Total sample	9,807	2,810
Overall absentee rate	23.1	13.5

[1] Absent less than two weeks due to illness, disability, or excused absence.
[2] Absent more than two weeks due to illness or disability.
[3] In school, but not invited to assessment session due to disruptive behavior.

Data Collection and Scoring

As with all NAEP assessments, Westat field staff conducted the data collection. Materials collected as part of the 1996 assessment were shipped to National Computer Systems (NCS), where NCS staff evaluated the responses to the constructed-response questions using scoring rubrics or guides prepared by the Educational Testing Service (ETS).

Each constructed-response question had a unique scoring guide that defined the criteria used to evaluate students' responses. The extended constructed-response questions were evaluated with four- or five-level guides (e.g., no evidence of understanding, evidence of minimal understanding, evidence of partial understanding, evidence of satisfactory understanding, or evidence of extended understanding). Short constructed-response questions were rated according to three-level guides (e.g., evidence of little or no understanding, evidence of partial understanding, and evidence of full understanding).

Student responses for constructed responses also were scored as "off task" if the student provided a response that was deemed not related in content to the question asked. A simple example of this type of response is, "I don't like this test." In contrast, responses scored as "incorrect" were valid attempts to answer the question that were simply wrong.

Data Analysis and IRT Scaling

Subsequent to the professional scoring, all information was transcribed to the NAEP database at ETS. Each processing activity was conducted with rigorous quality control.

Analyses were then conducted to determine the percentages of students who gave various responses to each cognitive and background question.

Item response theory (IRT) was used to estimate average scale scores. IRT models the probability of answering a question in a certain way as a mathematical function of proficiency or skill. The main purpose of IRT analysis is to provide a common scale, on which performance can be compared across groups—for example, those defined by characteristics such as gender and race/ethnicity.

Three distinct IRT models were used in scaling. Multiple-choice questions were scaled using the three-parameter logistic (3PL) model; short constructed-response questions rated as correct or incorrect were scaled using the two-parameter logistic (2PL) model; and short constructed-response questions rated according to a three-level rubric, as well as extended constructed-response questions rated on a four- or five-level rubric, were scaled using a generalized partial-credit model.

In the 1996 NAEP main science assessment, three distinct scales were created at each grade level to summarize students' abilities in the three defined fields of science: earth, physical, and life. All three scales ranged from 0 to 300 with a mean of 150 and a standard deviation of 35. The results from the advanced science study were reported using two scales: a biology scale and a chemistry/physics scale. These two scales were created by linking the results for students in the advanced science study to the performance of a comparable subgroup of students who took the main assessment (i.e., those students who were taking an advanced science course). The chemistry/physics scale was linked to the physical science scale in the main assessment, while biology was linked to the life science scale. The linking process was done via a linear transformation for each scale by matching the mean and standard deviation of all the plausible values based on students' performance in the advanced science study with the mean and standard deviation obtained for the corresponding scale in the main assessment. As a result, the mean chemistry/physics scale score is 175 with a standard deviation of 28, and the mean biology scale score is 173 with standard deviation of 28.

NAEP Reporting Groups

In this report, results are provided for three groups of students defined by shared characteristics—gender, race/ethnicity, and type of school. Results are reported for subpopulations only when sufficient numbers of students and adequate school representation are present. For public and nonpublic school students in the national assessment, the minimum requirement is at least 62 students in a particular subgroup from at least 5 primary sampling units (PSUs). However, the data for all students, regardless of whether their subgroup was reported separately, were included in computing overall results. Definitions of the subgroups referred to in this report are presented below.

Gender

Results are reported separately for males and females.

Race/ethnicity

The race/ethnicity variable is derived from school records and two questions asked of students from the set of general student background questions.

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

In the item above, 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 question, or provided information that was illegible or could not be classified, responses to the following question were examined to determine their race/ethnicity.

Which best describes you?

- ❍ White (not Hispanic)
- ❍ Black (not Hispanic)
- ❍ Hispanic (“Hispanic” means someone who is Mexican, Mexican American, Chicano, Puerto Rican, Cuban, or of 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 Islander background.)
- ❍ American Indian or Alaskan Native (“American Indian or Alaskan Native” means someone who is from one of the American Indian tribes or is one of the original people of Alaska.)
- ❍ Other (specify) ______________________________

Students’ race/ethnicity was then assigned on the basis of their responses. For students who filled in the sixth oval (“Other”), provided illegible information or information that could not be classified, or who did not respond at all, race/ethnicity was assigned as determined by school records.[11] Race/ethnicity could not be determined for students who did not respond to either of the demographic questions and whose schools did not provide information about race/ethnicity.

Details of how race/ethnicity classifications were derived are presented so that readers can determine how useful the results are for their particular purposes. Also, some students indicated that they were from a Hispanic background (e.g., Puerto Rican or Cuban) and that a racial/ethnic category other than Hispanic best describes them. These students were classified as Hispanic based on the rules described above. Furthermore, information from the schools did not always correspond to how students described themselves. Therefore, the racial/ethnic results presented in this report attempt to provide a clear picture based on several sources of information.

Type of school

Results are reported by the type of school that the student attended—public or nonpublic. Nonpublic schools include Catholic and other private schools. Although Bureau of Indian Affairs (BIA) schools and Department of Defense Domestic Dependent Elementary and Secondary Schools (DDESS) are not included in either the public or nonpublic categories, they are included in the overall national results.

[11] The procedure for assigning race/ethnicity was modified for Hawaii. See the *NAEP 1996 Technical Report* for details.

Estimating Variability

Because the statistics presented in this report are estimates of group and subgroup performance based on samples of students rather than the estimates that could be calculated if every student in the nation answered every question, the degree of uncertainty associated with the estimates should be taken into account. Two components of uncertainty are accounted for in the variability statistics based on student ability: (1) the uncertainty due to sampling only a relatively small number of students and (2) the uncertainty due to sampling only a relatively small number of cognitive questions. The first component accounts for the variability associated with the estimated percentages of students who had certain background characteristics or who answered a certain cognitive question correctly.

Because NAEP uses complex sampling procedures, conventional formulas for estimating sampling variability that assume simple random sampling are inappropriate. NAEP uses a jackknife replication procedure to estimate standard error. The jackknife standard error provides a reasonable measure of uncertainty for any student information that can be observed without error. However, because each student typically responds to only a few questions within any content strand, the scale score for any single student would be imprecise. In this case, plausible values technology can be used to describe the performance of groups or subgroups of students, but the underlying imprecision involved in this step adds another component of variability to statistics based on NAEP scale scores.[12]

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 error may be quite large. Throughout this report, estimates of standard errors subject to a large degree of uncertainty are designated.

The reader is reminded that, like findings from all surveys, NAEP results are subject to other kinds of error, including the effects of imperfect adjustments for student and school nonresponse and unknowable effects associated with the particular instrumentation and data collection methods. Nonsampling errors can be attributed to a number of sources—inability to obtain complete information about all selected schools in the sample (some students or schools refused to participate, or students participated but answered only certain questions); ambiguous definitions; differences in interpreting questions; inability or unwillingness to give correct information; mistakes in recording, coding, or scoring data; and other errors in collecting, processing, sampling, and estimating missing data. The extent of nonsampling error is difficult to estimate, and, because of their nature, the impact of such errors cannot be reflected in the data-based estimates of uncertainty provided in NAEP reports.

---

[12] For more details, see Johnson, E. G. & Rust, K. F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics, 17*(2), 175-190.

Drawing Inferences from the Results

As noted, because the percentages of students and their average scale scores are based on samples rather than on the entire population of twelfth-graders in the nation or a jurisdiction, the numbers reported are 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, the standard error should be taken into account, and observed similarities or differences should not be relied on solely. Therefore, the comparisons discussed in this report are based on statistical tests that consider the standard errors of those statistics and the magnitude of the difference among the averages or percentages.

The results from the sample, taking into account the uncertainty associated with all samples, are used to make inferences about the population. Using confidence intervals based on the standard errors provides a way to make inferences about the population averages 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 quantity. This statement means that one can conclude with approximately a 95 percent level of confidence that the average performance of the entire population of interest (e.g., all twelfth-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 group was 256, with a standard error of 1.2. A 95 percent confidence interval for the population quantity could be described in any of the following ways:

Average ±2 standard errors

$256 \pm 2 \times 1.2$

256 ± 2.4

253.6, 258.4

Thus, one can conclude with a 95 percent level of confidence that the average scale score for the entire population of students in that group is between 253.6 and 258.4.

Similar confidence intervals can be constructed for percentages, if the percentages are not extremely large or extremely small. For extreme percentages, confidence intervals constructed in the above manner may not be appropriate, and accurate confidence intervals can be constructed only 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 *NAEP 1996 Technical Report* contains a more complete discussion of extreme percentages.[13]

[13] Allen, N. L. Carlson, J. E., & Zelenak, C. A. (1999). The *NAEP 1996 technical report.* Washington, DC: National Center for Education Statistics.

Analyzing Group 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 percentages appear to be approximately the same. 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, regardless of whether the sample averages or percentages appear to be approximately the same or widely discrepant.

The reader is cautioned to rely on the results of the statistical tests rather than on the apparent magnitude of the difference between sample averages or percentages when determining whether the sample differences are likely to represent actual differences among the groups in the population.

To determine whether a real difference exists between the average scale scores (or percentages of a certain attribute) for two groups in the population, one needs to obtain an estimate of the degree of uncertainty associated with the difference between the averages (or percentages) of these groups 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 the squared standard errors, and taking the square root of that sum.

$$\text{Standard Error of the Difference} = SE_{A\text{-}B} = \sqrt{(SE_A^{\,2} + SE_B^{\,2})}$$

Similar to the way 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 among groups in the population are real. The difference between the averages or percentages 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 that a real difference between the groups is statistically significant (different) at the 0.05 level. In this report, differences among groups that involve poorly defined variability estimates or extreme percentages are not discussed.

As an example, to determine whether the average science scale score of group A is higher than that of group B, suppose that the sample estimates of the average scale score and standard errors were as follows:

Group	Average scale score	Standard error
A	218	0.9
B	216	1.1

The difference between the estimates of the average scale scores of groups A and B is 2 points (218–216). 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

Difference ± 2 standard errors of the difference

$2 \pm 2 \times 1.4$

2 ± 2.8

$-0.8, 4.8$

The value zero is within the confidence interval; therefore, there is insufficient evidence to claim that group A outperformed group B.

The procedures described in this section and the certainty ascribed to intervals (e.g., 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, many different groups are being compared (i.e., multiple sets of confidence intervals are being analyzed). 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. To hold the significance level for the set of comparisons at a particular level (e.g., 0.05), adjustments (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 obtain confidence intervals for the differences among groups when sets of comparisons were considered.[14] Thus, the confidence intervals for the sets of comparisons in the text are more conservative than those described on the previous pages.

Most of the multiple comparisons in this report pertain to relatively small sets or families of comparisons. In these situations, Bonferroni procedures were appropriate.

[14] Miller, R. G. (1966). *Simultaneous statistical inference.* New York: Wiley


$E=mc^2$

Appendix C

Standard Errors

Because NAEP scores and percentages are based on samples rather than the entire population(s), the results are subject to a measure of uncertainty reflected in the standard errors of the estimates. Standard errors provide a measure of how much results based on a sample would be expected to vary if a different, but equally valid, sample of students were chosen. This appendix contains the standard errors for the estimated averages and percentages in all the tables and figures throughout this report. It can be said with 95 percent certainty that for each population of interest, the value for the whole population is within plus or minus two standard errors of the estimate for the sample.

Table C2.1

Standard errors for students' reports on number of semesters of science taken from grades 9–12, by gender and by race/ethnicity: 1996

THE NATION'S REPORT CARD naep

From the beginning of ninth grade to the present, how many semesters of course work will you have taken in science?	Percentage of students				
	7 or more semesters	5-6 semesters	3-4 semesters	1-2 semesters	No semesters
Total					
Advanced study	2.3	1.5	1.3	0.2	***
Not eligible*	1.3	1.0	1.1	0.6	0.3
Male					
Advanced study	2.7	2.2	1.2	0.2	***
Not eligible*	1.6	1.2	1.4	1.1	0.4
Female					
Advanced study	2.9	1.8	1.8	0.3	***
Not eligible*	1.3	1.2	1.3	0.7	0.3
White					
Advanced study	2.3	1.6	1.2	0.1	***
Not eligible*	1.6	1.0	1.3	0.7	0.3
Black					
Advanced study	6.0	2.1	5.4	0.9	***
Not eligible*	1.8	1.7	2.2	1.8	0.6
Hispanic					
Advanced study	6.4	4.0	5.8	0.7	***
Not eligible*	2.4	1.9	2.5	2.4	0.9
Asian/Pacific Islander					
Advanced study	5.7	5.2	2.7	0.9	***
Not eligible*	3.1	3.7	3.2	2.5	***

*"Not eligible" indicates students who participated in main NAEP but were not eligible for the advanced science study.

***Sample size is insufficient to permit a reliable estimate.

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

Table C2.2

Standard errors for students' reports on amount of course work taken in biology from grades 9–12, by gender and race/ethnicity: 1996

THE NATION'S REPORT CARD naep

From the beginning of ninth grade to the present, how much science course work have you completed in biology?	Percentage of students			
	> 1 year	1 year	< 1 year	None
Total				
Advanced study	2.2	2.0	0.2	0.4
Not eligible*	0.9	1.2	0.3	0.6
Male				
Advanced study	2.4	2.2	0.2	0.6
Not eligible*	1.0	1.4	0.4	0.9
Female				
Advanced study	2.6	2.5	0.2	0.5
Not eligible*	1.1	1.4	0.6	0.5
White				
Advanced study	2.6	2.4	0.2	0.5
Not eligible*	1.0	1.4	0.4	0.7
Black				
Advanced study	2.7	2.8	***	0.4
Not eligible*	1.8	2.8	0.7	1.3
Hispanic				
Advanced study	4.4	4.9	***	1.0
Not eligible*	1.6	3.3	0.9	2.1
Asian/Pacific Islander				
Advanced study	5.5	5.7	0.4	***
Not eligible*	2.2	3.4	1.8	1.3

* "Not eligible" indicates students who participated in main NAEP but were not eligible for the advanced science study.

***Sample size is insufficient to permit a reliable estimate.

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

Table C2.3

Standard errors for students' reports on amount of course work taken in chemistry from grades 9–12, by gender and race/ethnicity: 1996

THE NATION'S REPORT CARD naep

From the beginning of ninth grade to the present, how much science course work have you completed in chemistry?	Percentage of students			
	> 1 Year	1 Year	< 1 Year	None
Total				
Advanced study	1.7	1.6	0.4	0.7
Not eligible*	0.5	1.6	0.5	1.6
Male				
Advanced study	1.7	1.7	0.5	0.9
Not eligible*	0.6	1.9	0.6	1.9
Female				
Advanced study	2.0	2.0	0.5	0.7
Not eligible*	0.5	1.7	0.6	1.6
White				
Advanced study	1.9	1.8	0.4	0.7
Not eligible*	0.6	1.8	0.7	1.7
Black				
Advanced study	2.8	3.2	1.7	1.1
Not eligible*	1.3	3.5	1.3	3.3
Hispanic				
Advanced study	2.5	3.5	1.0	2.5
Not eligible*	0.9	3.2	1.0	3.1
Asian/Pacific Islander				
Advanced study	5.4	5.0	0.8	1.3
Not eligible*	1.3	3.0	1.9	3.0

*"Not eligible" indicates students who participated in main NAEP but were not eligible for the advanced science study.

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

Table C2.4

Standard errors for students' reports on amount of course work taken in physics from grades 9–12, by gender and race/ethnicity: 1996

THE NATION'S REPORT CARD naep

From the beginning of ninth grade to the present, how much science course work have you completed in physics?	Percentage of students			
	> 1 Year	1 Year	< 1 Year	None
Total				
Advanced study	1.3	1.9	1.0	1.7
Not eligible*	0.3	1.3	0.6	1.5
Male				
Advanced study	1.6	2.2	1.1	2.0
Not eligible*	0.6	1.5	0.7	1.7
Female				
Advanced study	1.2	2.4	1.2	2.1
Not eligible*	0.3	1.4	0.8	1.7
White				
Advanced study	1.2	2.3	1.1	2.0
Not eligible*	0.4	1.6	0.9	1.9
Black				
Advanced study	2.6	5.5	2.5	6.9
Not eligible*	0.6	2.4	0.9	2.2
Hispanic				
Advanced study	2.5	5.3	4.1	5.2
Not eligible*	0.9	1.7	1.0	2.1
Asian/Pacific Islander				
Advanced study	4.0	3.8	2.0	2.5
Not eligible*	2.9	3.1	1.9	4.0

*"Not eligible" indicates students who participated in main NAEP but were not eligible for the advanced science study.

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

Table C2.5

Standard errors for students' reports on enrollment in Advanced Placement courses in science, by gender and race/ethnicity: 1996

THE NATION'S REPORT CARD naep

Are you currently enrolled in an Advanced Placement course in science?	Yes	No
Total		
Advanced study	2.6	2.6
Male		
Advanced study	2.5	2.5
Female		
Advanced study	3.1	3.1
White		
Advanced study	2.9	2.9
Black		
Advanced study	5.1	5.1
Hispanic		
Advanced study	4.8	4.8
Asian American/Pacific Islander		
Advanced study	6.1	6.1

NOTE: All students currently taking AP courses would have been eligible for the advanced science study—therefore, "Not eligible" results are not given.

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

Table C2.6 Standard errors for biology advanced science study scale scores, by gender, race/ethnicity, and type of school: 1996

THE NATION'S REPORT CARD naep

	Biology
Total	**1.5**
Male	1.6
Female	1.8
White	1.5
Black	3.1
Hispanic	3.4
Asian/Pacific Islander	3.4
Public schools	1.8
Nonpublic schools	2.7
Catholic schools	3.6
Other nonpublic	6.3

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

Table C2.7 Standard errors for chemistry/physics advanced science study scale scores by gender, race/ethnicity, and type of school: 1996

THE NATION'S REPORT CARD naep

	Chemistry/ physics
Total	**1.3**
Male	1.4
Female	1.4
White	1.1
Black	2.4
Hispanic	3.3
Asian/Pacific Islander	3.3
Public schools	1.5
Nonpublic schools	2.8
Catholic schools	4.4
Other nonpublic	6.1

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

Table C3.2 — *Standard errors for plant adaptation in desert: Percentage choosing each response: 1996*

THE NATION'S REPORT CARD naep

A	B	C	D	Omit
0.5	0.7	1.0	0.6	0.1

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

Table C3.3 — *Stanard errors for function of vaccine: Percentage at each score level: 1996*

THE NATION'S REPORT CARD naep

Unsatisfactory	*Partial*	*Complete*	Omit
1.2	1.3	1.3	0.7

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

Table C3.4

Standard errors for genetic counseling: Percentage at each score level: 1996

THE NATION'S REPORT CARD naep

Unsatisfactory	*Partial*	*Complete*	Omit
1.3	0.9	1.1	0.5

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

Table C3.5

Standard errors for volume to neutralize NaOH: Percentage choosing each response: 1996

THE NATION'S REPORT CARD naep

A	B	C	D	Omit
0.9	0.9	1.4	0.5	0.3

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

Table C3.6

Standard errors for grams used in reaction: Percentage choosing each response: 1996

THE NATION'S REPORT CARD naep

A	B	C	D	Omit
0.5	0.9	0.4	0.4	0.1

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

Figure C3.1 *Standard errors for advanced science study and main NAEP—Grams used in reaction: Percentage choosing correct response: 1996*

Advanced science study	Main NAEP
0.9	1.4

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

Table C3.7 *Standard errors for water at standard temperature and pressure: Percentage at each score level: 1996*

Unsatisfactory	*Partial*	*Complete*	Omit
1.5	1.3	0.4	1.4

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

Table C3.8 *Standard errors for ionization energy: Percentage at each score level: 1996*

Unsatisfactory	*Partial*	*Essential*	*Complete*	Omit
1.4	1.3	0.6	0.2	0.7

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

Table C3.9

Standard errors for kinetic energy of block: Percentage choosing each response: 1996

THE NATION'S REPORT CARD naep

A	B	C	D	Omit
0.6	0.8	1.2	0.7	0.2

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

Table C3.10

Standard errors for path of car on ice: Percentage choosing each response: 1996

THE NATION'S REPORT CARD naep

A	B	C	D	Omit
0.4	0.9	1.3	0.7	0.1

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

Figure C3.2

Standard errors for Advanced science study and main NAEP—Path of car on ice: Percentage choosing the correct response: 1996

THE NATION'S REPORT CARD naep

Advanced science study	Main NAEP
1.3	2.3

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

Table C3.11

Standard errors for length of string: Percentage at each score level: 1996

Unsatisfactory	Partial	Complete	Omit
1.4	0.9	0.6	1.1

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

Table C3.12

Standard errors for direction and speed of train relative to car: Percentage at each score level: 1996

Unsatisfactory	Partial	Essential	Complete	Omit
1.4	0.9	0.9	1.0	***

***Sample size is insufficient to permit a reliable estimate.

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

Acknowledgments

This report is the culmination of the efforts of many individuals who contributed their considerable knowledge, expertise, 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, 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 thousands of students and hundreds of teachers and administrators who made the assessment possible.

The 1996 NAEP science assessment and advanced science study were funded through NCES, in the Office of Educational Research and Improvement of the U. S. Department of Education. The acting Commissioner of Education Statistics, Gary W. Phillips, and the NCES staff—Janis Brown, Peggy Carr, Pat Dabbs, Arnold Goldstein, Carol Johnson, Andrew Kolstad, Holly Spurlock, and Shi-Chang Wu—worked closely and collegially with the authors to produce this report.

The NAEP project at ETS is directed by Steven 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 advanced science 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, Mathilde Kennel, Linda Reynolds, and Brent Studer.

The statistical and psychometric activities for NAEP at ETS are directed by Nancy Allen, John Barone, John Donoghue, and David Freund. The analyses presented in this report were led by John Donoghue and Steven Isham, with assistance from Jinming Zhang, Lois Worthington, and Kate Pashley.

The considerable production of the report was overseen by Loretta Casalaina, Carol Errickson, and Joe Kolodey. Shari Santapau coordinated the editorial reviews and Alice Kass assisted with data checking. The production of the World Wide Web version of this report was led by Pat O'Reilly.

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: Ghedom Bairu, Mary Lyn Bourque, Jay Campbell, Angelo Collins, Patricia Dabbs, Lawrence Feinberg, Arnold Goldstein, Carol Johnson, Janet Johnson, Andrew Kolstad, Anthony Lutkus, Marilyn McMillen, Laurence Ogle, Will Pfeiffenberger, Taslima Rahmon, Senta Raizen, Holly Spurlock, and Alan Vanneman.

United States
Department of Education
ED Pubs
8242-B Sandy Court
Jessup, MD 20794-1398