

Appendix A

Technical Procedures for the NAEP 2011 Science Assessment

This appendix provides an overview of some of the technical procedures for the NAEP 2011 science assessment. The assessment was administered in 2011 at grade 8 so that results from both the NAEP mathematics and science assessments could be linked to results from the 2011 Trends in International Mathematics and Science Study (TIMSS). NAEP science was not administered at any other grades in 2011. Information is included about the content of the assessment, school and student samples and participation, inclusion of students with disabilities and/or English language learners, analysis procedures, and interpretation of results. Additional technical information about NAEP assessments is available on the Web at <http://nces.ed.gov/nationsreportcard/tdw/>.

Development of the Science Framework

The National Assessment Governing Board oversees the creation of the NAEP frameworks that describe the specific knowledge and skills that should be assessed in each subject. The frameworks also provide the theoretical basis for the assessment, direction for what types of items should be included, and how the items should be designed and scored. While the frameworks describe the general content and design of NAEP subject area assessments, the specifications provide the detailed information used by test developers for constructing the assessments and more detailed information in scoring. Both the *Science Framework for the 2011 National Assessment of Educational Progress* and item specifications are available on the Governing Board's website at <http://www.nagb.org/publications/frameworks.htm>.

The 2009 NAEP science framework approved by the Governing Board replaced the framework used for the 1996, 2000, and 2005 science assessments. A variety of factors made it necessary to create a new framework to guide the assessment of science in 2009 and beyond: the publication of *National Standards* for science literacy, advances in both science and cognitive research, the growth in the prevalence of national and international science assessments, advances in innovative assessment approaches, and the need to advance the state of the art so that the widest possible range of students can be fairly assessed. The framework is unchanged for 2011.

The development of the new science framework involved the critical input of hundreds of individuals across the country, including some of the nation's leading scientists, science educators, policymakers, and assessment experts. Under contract to the Governing Board, WestEd and the Council of Chief State School Officers (CCSSO) spent 18 months developing the framework; this process involved committees, regional hearings, and other public forums. The Governing Board also engaged an external review panel to evaluate the draft framework and convened a public hearing to receive additional input during the development process.

The frameworks for all main NAEP assessments are periodically updated or changed to reflect current curricula and standards. Whenever changes are made to a subject framework, every effort is made to maintain the trend lines that permit the reporting of changes in student achievement over time. If, however, the changes made to an assessment are such that the results are not comparable to earlier assessments, a new trend line is started. The assessment resulting from the 2009 framework started a new NAEP science trend.

Framework Dimensions

The design of the NAEP science assessment is guided by the framework's descriptions of the science content and practices to be assessed. Students are expected to have learned science content comprised of the facts, concepts, laws, principles, and theories that have been verified by the community of scientists, as well as understand how scientists gather, organize, and evaluate empirical evidence. Each question in the 2011 science assessment was classified based on two dimensions: *science content* and *science practices*. By considering these two dimensions for each question, the framework ensures that NAEP assesses an appropriate balance of content along with a variety of ways of doing science.

SCIENCE CONTENT

The 2011 framework organizes science content into three broad content areas reflecting the science curriculum students are generally exposed to across the K–12 curriculum, including physical science, life science, and Earth and space sciences.

- Physical science includes concepts related to properties and changes of matter, forms of energy, energy transfer and conservation, position and motion of objects, and forces affecting motion.
- Life science includes concepts related to organization and development, matter and energy transformations, interdependence, heredity and reproduction, and evolution and diversity.
- Earth and space sciences include concepts related to objects in the universe, the history of the Earth, properties of Earth materials, tectonics, energy in Earth systems, climate and weather, and biogeochemical cycles.

SCIENCE PRACTICES

In addition to the science content, the framework assesses student understanding of how scientific knowledge is used by measuring what students are able to do with the science content. Four science practices describe how science knowledge is used—identifying science principles, using science principles, using scientific inquiry, and using technological design.

- Identifying science principles focuses on students' ability to recognize, recall, define, relate, and represent basic science principles in each of the three content areas.
- Using science principles focuses on the importance of science knowledge in making accurate predictions about and explaining observations of the natural world.
- Using scientific inquiry focuses on designing, critiquing, and evaluating scientific investigations; identifying patterns in data; and using empirical evidence to validate or criticize conclusions.
- Using technological design focuses on the systematic process of applying science knowledge and skills to propose or critique solutions to real world problems, identify trade-offs, and anticipate effects of technological design decisions.

The distribution of items across the four science practices is as follows: Identifying Science Principles and Using Science Principles (combined), 60 percent; Using Scientific Inquiry, 30 percent; and Using Technological Design, 10 percent.

Table A-1. Percentage distribution of target and actual assessment time in NAEP science at grade 8, by field of science: 2011

Content area	Target	Actual
Physical science	30	26
Life science	30	34
Earth and space sciences	40	41

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Science Assessment.

Content of the 2011 Science Assessment

Each NAEP assessment contains two major components: subject-specific cognitive items that measure the achievement of students in an academic subject; and background items that collect information from students, teachers, and school administrators about variables that are related to student achievement. Both the cognitive and background items are developed through a process that includes reviews by external advisory groups and field testing. Results from the cognitive items provide information about what students know and can do in a subject area. Information from the background items gives context to NAEP results and allows researchers to track factors associated with academic achievement.

The 2011 science assessment was made up of 144 questions at eighth grade. Students spent about one-half of the assessment time responding to multiple-choice questions and one-half responding to two types of constructed-response questions. Short constructed-response questions required students to write a concise explanation for a given situation or result, illustrate with a brief example, or describe a quantitative relationship in response to the question provided. Extended constructed-response questions were generally multidimensional and required students to solve a problem by applying and integrating science concepts and required that students analyze a science situation and explain a concept. Table A-2 shows the number of cognitive items administered in 2011 by item format.

Table A-2. Number of NAEP science questions at grade 8, by question type: 2011

Question type	Number of questions
Total	144
Multiple-choice	94
Short constructed response	30
Extended constructed response	20

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Science Assessment.

Cognitive Blocks: The assessment design allowed for broad coverage at each grade of the three science content areas and four science practices, while minimizing the time burden for any one student. This was accomplished through the use of matrix sampling of items in which each student was required to take only a small portion of the entire pool of assessment questions.

The science item pool was organized into subsets or "blocks." In 2011, there were a total of 9 blocks at eighth grade. Each science assessment booklet contained two separately timed 25-minute blocks. Each block contained between 14 and 18 questions, depending on the balance between multiple-choice and constructed-response questions.

The procedure used to create booklets ensured that each block was paired with every other block. In addition, the procedure controlled for possible block-position effects across the set of booklets by balancing the order of the blocks within booklets. The booklets were cycled through in such a way that each booklet was used approximately an equal number of times across the entire assessment, while no more than a few students in any given assessment session received the same booklet.

Sample released questions at all three grade levels can be viewed at the NAEP website at <http://nces.ed.gov/nationsreportcard/itmrlsx/>. Items may be sorted by difficulty and question type.

NAEP Samples

NAEP assesses representative samples of students rather than the entire population of students. The sample selection process utilizes a probability sample design in which each school and each student has a known probability of being selected (the probabilities are proportionate to the estimated number of students in the grade of an assessed school). Samples are selected according to a multistage design, with students drawn from within sampled public and private schools nationwide.

The 2007-08 Common Core of Data (CCD) file, a comprehensive list of operating public schools in each jurisdiction that is compiled each school year by the National Center for Education Statistics, served as the sampling frame for the selection of public schools in each state/jurisdiction. All students at more local geographic sampling levels also make up part of the broader samples. For example, the state samples are included as part of the national sample.

The 2007-08 Private School Survey (PSS), a mail survey of all U.S. private schools carried out biennially by the Census Bureau under contract to NCES, served as the sampling frame for private schools. While state and district results are based on samples of public schools only, the national results are based on the combined samples of public and private schools. Although information about the combined public and private school national samples is provided here for context, performance results in the State Report Generator are for public school students only.

Table A-3 shows the target populations and sample sizes in 2011 for the nation and participating states and jurisdictions at grade 8.

Because each school that participated in the assessment, and each student assessed, represents only a portion of the larger population of interest, the results are weighted to make appropriate inferences between the student samples and the respective populations from which they are drawn. Sampling weights are adjusted for the disproportionate representation of some groups in the selected sample. This includes oversampling of schools with high concentrations of students from certain racial/ethnic groups and the lower sampling rates of students who attend very small schools.

Table A-3. Student sample size and target population in NAEP science at grade 8, by state/jurisdiction: 2011

State/jurisdiction	Sample size	Target population
Nation	124,200	3,821,000
Public	121,800	3,508,000
Private	800	306,000
Alabama	2,300	55,000
Alaska	2,100	8,000
Arizona	2,300	75,000
Arkansas	2,400	35,000
California	2,500	462,000
Colorado	1,900	55,000
Connecticut	2,200	40,000
Delaware	2,300	9,000
Florida	2,300	191,000
Georgia	2,400	112,000
Hawaii	2,400	12,000
Idaho	2,400	19,000
Illinois	3,600	147,000
Indiana	2,300	75,000
Iowa	2,200	33,000
Kansas	2,400	32,000
Kentucky	3,300	49,000
Louisiana	2,200	46,000
Maine	2,300	14,000
Maryland	2,300	61,000
Massachusetts	2,400	73,000
Michigan	2,300	114,000
Minnesota	2,500	59,000
Mississippi	2,200	35,000
Missouri	2,200	60,000
Montana	2,200	10,000
Nebraska	2,200	20,000
Nevada	2,300	32,000
New Hampshire	2,300	15,000
New Jersey	2,200	95,000
New Mexico	2,900	23,000
New York	3,500	208,000
North Carolina	2,600	103,000
North Dakota	2,000	7,000
Ohio	2,300	125,000
Oklahoma	2,100	41,000
Oregon	2,400	42,000
Pennsylvania	2,300	138,000
Rhode Island	2,300	11,000
South Carolina	2,400	51,000
South Dakota	2,600	9,000
Tennessee	2,400	68,000
Texas	2,800	341,000
Utah	2,500	38,000
Vermont	1,800	6,000
Virginia	2,400	86,000
Washington	2,700	78,000
West Virginia	2,400	19,000
Wisconsin	2,200	58,000
Wyoming	1,800	6,000
Other jurisdictions		
BIE ¹	100	2,000
District of Columbia	2,500	4,000
DoDEA ²	1,400	5,000

¹ Bureau of Indian Education.

² Department of Defense Education Activity (overseas and domestic schools).

NOTE: The sample size is rounded to the nearest hundred. The target population is rounded to the nearest thousand. Data for BIE and DoDEA schools are counted in the overall nation total, but not in the nation (public) total. Data for the District of Columbia public schools are counted, along with the states, in nation (public). Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Science Assessment.

School and Student Participation

National Participation

To ensure unbiased samples, the National Assessment Governing Board policy on reporting requires that weighted participation rates for original school samples be 70 percent or higher, for public and private schools respectively, to report national results separately for public and private schools. In instances where the original weighted school participation rate falls below 85 percent, NCES statistical standards require that a nonresponse bias analysis be conducted to determine if the responding school sample is not representative of the population, thereby introducing the potential for nonresponse bias. The decision whether or not to report the results in a case where the response rate falls between 70 and 85 percent depends upon the results of this nonresponse bias analysis.

National school and student participation rates for the 2011 science assessment are presented in table A-4. Student-weighted school participation rates were 97 percent for grade 8 (100 percent for public schools and 74 percent for private schools). Weighted student participation rates were 93 percent for grade 8 (93 percent for public schools and 94 percent for private schools).

Table A-4. National school and student participation rates in NAEP science at grade 8, by type of school: 2011

Type of school	School participation					Student participation	
	Student weighted		School weighted		Number of schools participating after substitution	Student-weighted percent	Number of students assessed
	Percent before substitution	Percent after substitution	Percent before substitution	Percent after substitution			
Nation	97	98	88	92	7,290	93	122,000
Public	100	100	100	100	6,690	93	119,600
Private	74	85	70	80	480	94	800

NOTE: The national totals for schools include Department of Defense Education Activity (overseas and domestic schools) and Bureau of Indian Education schools, which are not included in either the public or private totals. The national totals for students include students in these schools. Columns of percentages have different denominators. The number of schools is rounded to the nearest ten. The number of students is rounded to the nearest hundred.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Science Assessment.

The student-weighted school participation rates are calculated based on school sampling weights and grade-specific school enrollment figures. The denominator of the rate is the weighted total number of students represented by the initially selected schools that had eligible students enrolled. This includes both participating and nonparticipating schools. The numerator is the weighted total number of students represented by participating schools. This is calculated in two distinct ways: first, with participating schools defined as only the initially selected schools that participated in the assessment (which gives rise to the rate before substitution), and second, with all of the participating schools after substitution (giving the rate after substitution). On the other hand, the school-weighted school participation rates are calculated based only on the school sampling weights. They show the weighted total number of schools (either before or after substitution) divided by the weighted total number of schools in the initially selected sample.

State Participation

Standards established by the Governing Board require that student-weighted school participation rates for the state samples need to be at least 85 percent for results to be reported. In 2011, fifty-one states and jurisdictions participating in the science assessment at grade 8 met this participation rate requirement, with the exception of Colorado, where the participation rate was 84 percent (table A-5). Note that no school substitution was used for the state samples at grade 8.

Table A-5. Public school and student participation rates in NAEP science at grade 8, by state/jurisdiction: 2011

State/jurisdiction	School participation			Student participation	
	Student-weighted percent	School-weighted percent	Number of schools participating	Student-weighted percent	Number of students assessed
Nation (public)	100	100	6,690	93	119,600
Alabama	100	100	110	93	2,300
Alaska	100	98	120	90	2,100
Arizona	99	99	120	93	2,300
Arkansas	100	100	120	94	2,300
California	100	100	220	93	2,500
Colorado	84	87	100	93	1,900
Connecticut	100	100	110	91	2,200
Delaware	100	100	50	92	2,300
Florida	100	100	210	93	2,300
Georgia	100	100	120	93	2,400
Hawaii	100	100	80	93	2,400
Idaho	100	100	110	93	2,400
Illinois	100	100	210	94	3,500
Indiana	100	100	110	94	2,300
Iowa	100	100	130	93	2,200
Kansas	100	100	140	94	2,300
Kentucky	100	100	140	93	3,200
Louisiana	100	100	120	93	2,200
Maine	100	100	130	93	2,200
Maryland	99	99	150	93	2,300
Massachusetts	99	98	140	92	2,300
Michigan	100	100	150	92	2,200
Minnesota	100	100	140	92	2,500
Mississippi	100	100	110	92	2,100
Missouri	100	100	120	93	2,100
Montana	100	98	180	91	2,200
Nebraska	100	100	140	95	2,200
Nevada	100	97	90	93	2,300
New Hampshire	100	100	90	91	2,200
New Jersey	100	100	110	92	2,200
New Mexico	99	99	120	92	2,800
New York	99	100	170	91	3,400
North Carolina	100	100	150	92	2,600
North Dakota	100	99	170	95	1,900
Ohio	100	100	160	93	2,300
Oklahoma	100	100	150	92	2,100
Oregon	99	99	140	93	2,400
Pennsylvania	100	100	160	93	2,300
Rhode Island	100	100	50	92	2,300
South Carolina	100	100	110	94	2,300
South Dakota	100	100	210	95	2,600
Tennessee	100	100	120	92	2,400
Texas	99	100	210	93	2,700
Utah	100	100	110	92	2,400
Vermont	100	100	120	94	1,800
Virginia	100	100	110	94	2,300
Washington	100	100	130	92	2,600
West Virginia	100	100	110	93	2,300
Wisconsin	100	100	150	93	2,100
Wyoming	100	100	90	92	1,800
Other jurisdictions					
District of Columbia	100	100	80	88	2,500
DoDEA ¹	99	95	60	94	1,400

¹ Department of Defense Education Activity (overseas and domestic schools).

NOTE: The number of schools is rounded to the nearest ten. The number of students is rounded to the nearest hundred. The school participation rates are student-weighted percentages before substitution. Columns of percentages have different denominators. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Science Assessment.

Inclusion of Students With Disabilities and/or English Language Learners

It is important for NAEP to assess as many students selected to participate as possible. Assessing representative samples of students, including students with disabilities (SD) and English language learners (ELL), helps to ensure that NAEP results accurately reflect the educational performance of all students in the target population, and can continue to serve as a meaningful measure of U.S. students' academic achievement over time.

The National Assessment Governing Board, which sets policy for NAEP, has been exploring ways to ensure that NAEP continues to appropriately include as many students as possible and to do so in a consistent manner for all jurisdictions assessed and reported. In March 2010, the Governing Board adopted a new policy, NAEP Testing and Reporting on Students with Disabilities and English Language Learners. This policy was the culmination of work with experts in testing and curriculum, and those who work with exceptional children and students learning to speak English. The policy aims to

- maximize participation of sampled students in NAEP,
- reduce variation in exclusion rates for SD and ELL students across states and districts,
- develop uniform national rules for including students in NAEP, and
- ensure that NAEP is fully representative of SD and ELL students.

The policy defines specific inclusion goals for NAEP samples. At the national, state, and district levels, the goal is to include 95 percent of all students selected for the NAEP samples, and 85 percent of those in the NAEP sample who are identified as SD or ELL.

Students are selected to participate in NAEP based on a sampling procedure designed to yield a sample of students that is representative of students in all schools nationwide and in public schools within each state. First, schools are selected, and then students are sampled from within those schools without regard to disability or English language proficiency. Once students are selected, those previously identified as SD or ELL may be offered accommodations or excluded.

States and jurisdictions vary in their proportions of special-needs students and in their policies on inclusion and the use of accommodations. Despite the increasing identification of SD and ELL students in some states, in particular of ELL students at grade 4, NAEP inclusion rates have generally remained steady or increased since 2003. Only a small number of states included a smaller percentage of students in the 2011 NAEP science assessments than in 2009. Inclusion rates decreased by 1 percentage point for only 1 jurisdiction. This reflects efforts on the part of states and jurisdictions to include all students who can meaningfully participate in the NAEP assessments, as well as the historically high inclusion rates for science. The new NAEP inclusion policy is an effort to ensure that this trend continues.

Determining whether each jurisdiction has met the NAEP inclusion goals involves looking at three different inclusion rates—an overall inclusion rate, an inclusion rate for SD students, and an inclusion rate for ELL students. Each inclusion rate is calculated as the percentage of sampled students who were included in the assessment (i.e., were not excluded).

Inclusion rate percentages are estimates because they are based on representative samples of students rather than on the entire population of students. As such, the inclusion rates are associated with a margin of error. The margin of error for each jurisdiction's inclusion rate was taken into account when comparing it to the corresponding inclusion goal. For example, if the point estimate of a state's overall inclusion rate was 93 percent and had a margin of error of plus or minus 3 percentage points, the state was considered to have met the 95 percent inclusion goal because the 95 percent goal falls within the margin of error, which ranges from 90 percent to 96 percent. Refer to the Technical Notes for more details about how the margin of error was used in these calculations.

Variations in inclusion rates across jurisdictions or from year to year may affect the comparability of results. Because SD and ELL students tend to score lower than average, it might be expected that excluding more of these students would tend to raise scores and that including more would tend to lower scores. However, across states, correlations between inclusion rates and average 2011 science scores at grade 8 (.03) showed only a weak association. With regard to state trends, changes in the percentages of students included and changes in average science scores from 2009 to 2011 showed a weak negative association at grade 8 (-.11). Therefore, there was a weak tendency at grade 8 for states with score gains to also have excluded a larger percentage of students in 2011 compared to 2009.

Confidence intervals for state inclusion rates

NAEP endeavors to include as many sampled students as possible in the assessment, including students with disabilities (SD) and English language learners (ELL), and has established specific inclusion goals: 95 percent of all sampled students and 85 percent of sampled students identified as SD or ELL. Inclusion rates were computed for each state/jurisdiction participating in the 2011 assessment and compared to NAEP inclusion goals. Three inclusion percentages were computed for each state/jurisdiction. An overall inclusion percentage represents included students as a percentage of all students sampled within the state/jurisdiction. In addition, separate percentages were computed to report included students as a percentage of the state/jurisdiction sample that was identified as SD or ELL.

Inclusion percentages are estimates based on a sample, and each estimate has a measure of uncertainty or margin of error. Confidence intervals quantify this uncertainty due to sampling, resulting in interval estimates of the inclusion percentages. Therefore, confidence intervals for inclusion percentages were used to determine upper and lower confidence bounds around the inclusion point estimates.

When determining whether each state/jurisdiction met the NAEP inclusion goals, the confidence intervals were used, rather than just the point estimates. This means that if the inclusion goal of either 95 percent or 85 percent fell within the corresponding confidence interval, the state/jurisdiction was considered as having met the goal. States/jurisdictions for which the upper bound of the confidence interval was less than 95 percent (or 85 percent) did not meet the inclusion goal.

See the National Assessment Governing Board's policy on NAEP Testing and Reporting on Students with Disabilities and English Language Learners at [http://www.nagb.org/policies/PoliciesPDFs/Reporting and Dissemination/naep_testandreport_studentswithdisabilities.pdf](http://www.nagb.org/policies/PoliciesPDFs/Reporting%20and%20Dissemination/naep_testandreport_studentswithdisabilities.pdf).

All of the states/jurisdictions participating in the 2011 science assessment met the 95 percent inclusion goal. See appendix table A-6 for the inclusion rates as a percentage of all students selected in each state/jurisdiction, and table A-7 for the rates as a percentage of the SD or ELL students.

Table A-6. Inclusion rate and confidence interval in NAEP science for eighth-grade public school students, as a percentage of all students, by state/jurisdiction: 2011

State/jurisdiction	Inclusion rate	95% confidence interval	
		Lower	Upper
Nation (public)	98¹	98.2	98.4
Alabama	99 ¹	98.4	99.3
Alaska	99 ¹	98.5	99.2
Arizona	99 ¹	98.4	99.5
Arkansas	99 ¹	98.5	99.4
California	98 ¹	97.6	98.7
Colorado	99 ¹	98.5	99.4
Connecticut	99 ¹	98.1	99.1
Delaware	98 ¹	97.7	98.8
Florida	99 ¹	98.3	99.1
Georgia	98 ¹	97.7	98.9
Hawaii	98 ¹	97.3	98.5
Idaho	99 ¹	98.0	98.9
Illinois	99 ¹	98.4	99.2
Indiana	99 ¹	97.9	99.2
Iowa	99 ¹	98.4	99.4
Kansas	99 ¹	98.0	99.0
Kentucky	97 ¹	96.5	97.9
Louisiana	99 ¹	98.3	99.2
Maine	98 ¹	97.5	98.7
Maryland	98 ¹	97.4	98.6
Massachusetts	97 ¹	95.8	97.6
Michigan	97 ¹	96.5	97.9
Minnesota	98 ¹	97.2	98.7
Mississippi	99 ¹	98.7	99.4
Missouri	99 ¹	98.1	99.2
Montana	98 ¹	97.8	98.9
Nebraska	99 ¹	98.0	99.0
Nevada	99 ¹	98.0	99.2
New Hampshire	98 ¹	97.2	98.4
New Jersey	99 ¹	98.1	99.2
New Mexico	98 ¹	97.7	98.7
New York	99 ¹	98.0	99.1
North Carolina	98 ¹	97.8	98.8
North Dakota	97 ¹	95.9	97.5
Ohio	98 ¹	97.0	98.5
Oklahoma	97 ¹	96.1	97.9
Oregon	98 ¹	97.7	98.9
Pennsylvania	99 ¹	98.3	99.4
Rhode Island	99 ¹	99.0	99.6
South Carolina	99 ¹	98.2	99.2
South Dakota	99 ¹	98.3	99.1
Tennessee	99 ¹	97.7	99.1
Texas	98 ¹	96.8	98.3
Utah	98 ¹	97.6	98.6
Vermont	99 ¹	98.1	99.0
Virginia	97 ¹	96.1	98.2
Washington	98 ¹	97.4	98.7
West Virginia	98 ¹	97.8	98.8
Wisconsin	98 ¹	97.2	98.7
Wyoming	99 ¹	98.1	99.1
Other jurisdictions			
District of Columbia	99 ¹	98.0	98.9
DoDEA ²	99 ¹	98.1	99.1

¹ The state/jurisdiction's inclusion rate is higher than or not significantly different from the National Assessment Governing Board goal of 95 percent.

² Department of Defense Education Activity (overseas and domestic schools).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Science Assessment.

Table A-7. Inclusion rate and standard error (SE) in NAEP science for eighth-grade public school students with disabilities (SD) and English language learners (ELL), as a percentage of identified SD or ELL students, by state/jurisdiction: 2011

State/jurisdiction	Percentage of identified SD or ELL students			
	SD		ELL	
	Inclusion rate	SE	Inclusion rate	SE
Nation (public)	87¹	0.5	94¹	0.5
Alabama	90 ¹	2.2	‡	†
Alaska	92 ¹	1.4	97 ¹	1.1
Arizona	92 ¹	2.4	‡	†
Arkansas	91 ¹	2.1	98 ¹	1.3
California	82 ¹	2.5	96 ¹	0.8
Colorado	90 ¹	2.1	98 ¹	1.0
Connecticut	88 ¹	2.1	93 ¹	2.9
Delaware	89 ¹	1.8	‡	†
Florida	93 ¹	1.3	93 ¹	2.6
Georgia	85 ¹	2.9	‡	†
Hawaii	91 ¹	1.9	89 ¹	1.9
Idaho	83 ¹	3.0	94 ¹	2.5
Illinois	93 ¹	1.3	94 ¹	2.3
Indiana	91 ¹	2.1	98 ¹	1.5
Iowa	94 ¹	1.7	98 ¹	2.3
Kansas	89 ¹	2.3	98 ¹	0.9
Kentucky	79	2.6	‡	†
Louisiana	89 ¹	1.7	‡	†
Maine	90 ¹	1.5	‡	†
Maryland	85 ¹	2.7	‡	†
Massachusetts	84 ¹	2.2	83 ¹	3.9
Michigan	80	2.7	‡	†
Minnesota	86 ¹	2.4	95 ¹	2.3
Mississippi	89 ¹	2.3	‡	†
Missouri	90 ¹	2.1	‡	†
Montana	87 ¹	2.2	‡	†
Nebraska	91 ¹	1.7	94 ¹	2.5
Nevada	89 ¹	2.4	96 ¹	1.4
New Hampshire	88 ¹	1.6	‡	†
New Jersey	93 ¹	1.4	‡	†
New Mexico	87 ¹	1.9	93 ¹	1.4
New York	93 ¹	1.4	92 ¹	2.5
North Carolina	88 ¹	2.1	95 ¹	1.9
North Dakota	76	2.6	‡	†
Ohio	85 ¹	2.5	‡	†
Oklahoma	83 ¹	2.6	‡	†
Oregon	88 ¹	2.1	96 ¹	1.9
Pennsylvania	93 ¹	1.5	‡	†
Rhode Island	97 ¹	0.8	92 ¹	3.1
South Carolina	90 ¹	2.1	99 ¹	1.0
South Dakota	90 ¹	1.7	‡	†
Tennessee	87 ¹	2.9	‡	†
Texas	80 ¹	3.0	91 ¹	2.5
Utah	83 ¹	2.2	95 ¹	1.8
Vermont	92 ¹	1.3	‡	†
Virginia	85 ¹	2.6	83 ¹	5.5
Washington	85 ¹	2.4	96 ¹	2.4
West Virginia	88 ¹	1.9	‡	†
Wisconsin	87 ¹	2.4	97 ¹	1.8
Wyoming	90 ¹	2.0	‡	†
Other jurisdictions				
District of Columbia	93 ¹	1.2	91 ¹	2.0
DoDEA ²	94 ¹	1.8	83 ¹	4.3

† Not applicable. Standard error estimate cannot be accurately determined.

‡ Reporting standards not met. Sample size insufficient to permit a reliable estimate.

¹ The state/jurisdiction's inclusion rate is higher than or not significantly different from the National Assessment Governing Board goal of 85 percent.

² Department of Defense Education Activity (overseas and domestic schools).

NOTE: SD includes students identified as having a Individualized Education Program but excludes other students protected under section 504 of the Rehabilitation Act of 1973.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Science Assessment.

Accommodations

Prior to 1996, no testing accommodations were provided to students taking the NAEP assessments, resulting in the exclusion of students who could not be assessed without them. As the number of identified students with disabilities and English language learners increased over the years, the exclusion of those needing accommodations to participate in NAEP threatened the stability of trend lines (excluding more students in one assessment year than in another might lead to apparent rather than real differences), and threatened to compromise NAEP samples as optimally representative of target populations. Therefore, administration procedures allowing for many of the same testing accommodations provided on state and district assessments (e.g., extra testing time or individual rather than group administration) were introduced in 1996 for national NAEP assessments and in 2000 for NAEP state assessments.

The percentages of SD/ELL students assessed with the available accommodations in 2011 are presented in table A-8. Students assessed with accommodations typically received some combination of accommodations. For example, students assessed in small groups (as compared with standard NAEP sessions of about 30 students) were also usually given extended time and are included in counts for both groups in table A-8.

Table A-8. Percentage of eighth-grade public and nonpublic school students identified as students with disabilities (SD) and/or English language learners (ELL) assessed in NAEP science with accommodations, by SD/ELL category and type of accommodation: 2011

Type of accommodation	SD and/or ELL	SD	ELL
Bilingual booklet	0.1	#	0.1
Bilingual dictionary	0.6	0.1	0.6
Braille version of the text	#	#	#
Breaks	1.5	1.4	0.2
Cue to stay on task	0.5	0.5	0.1
Directions read aloud in English	2.1	1.9	0.4
Directions read aloud in Spanish	0.1	#	0.1
Extended time	8.7	7.5	1.7
Large-print booklet	#	#	#
Magnification device	#	#	#
One-on-one	0.4	0.3	#
Read aloud (all or most of assessment)	3.5	3.3	0.5
Read aloud (occasional words or phrases)	0.9	0.8	0.2
Read aloud in Spanish (all or most of assessment)	#	#	#
School staff administers	0.3	0.3	#
Scribe	0.2	0.2	#
Sign language	#	#	#
Small group	7.7	7.0	1.3
Special equipment	0.3	0.3	#
Other	0.2	0.2	#

Rounds to zero.

NOTE: Students identified as both SD and ELL were counted only once under the combined SD and/or ELL category, but were counted separately under the SD and ELL categories.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Science Assessment.

Exclusion Rates

Even with the availability of accommodations, some students are excluded from the NAEP assessments by their schools. The decision to exclude any student is made by school staff who, using NAEP guidelines and each student's Individualized Education Program (IEP), decide whether the student can meaningfully be assessed.

Jurisdictions vary in their proportions of special-needs students. These variations, as well as differences in policies and practices regarding the identification and inclusion of special-needs students, lead to differences in exclusion and accommodation rates. These differences should be considered when comparing student performance over time and across jurisdictions. While the effect of exclusion is not precisely known, the validity of comparisons of performance results could be affected if exclusion rates are comparatively high or vary widely over time.

National Exclusion Rates (public and nonpublic school students): In 2011, seventeen percent at grade eight were identified as SD and/or ELL, with 2 percent excluded at grade eight (table A-9). The percentage of SD and/or ELL students assessed with accommodations in 2011 was 11 percent at grade eight. The proportions of SD and/or ELL students excluded and assessed with and without accommodations as a percentage of students identified are provided in table A-10.

Table A-9. Percentage of eighth-grade public and nonpublic school students with disabilities (SD) and/or English language learners (ELL) identified, excluded, and assessed in NAEP science: 2009 and 2011

SD/ELL category	2009	2011
SD and/or ELL		
Identified	17	17
Excluded	2	2
Assessed	15	15
Without accommodations	5	4
With accommodations	10	11
SD		
Identified	12	12
Excluded	2	1
Assessed	11	11
Without accommodations	2	2
With accommodations	9	9
ELL		
Identified	5	6
Excluded	#	#
Assessed	5	5
Without accommodations	3	3
With accommodations	2	2

Rounds to zero.

NOTE: Students identified as both SD and ELL were counted only once under the combined SD and/or ELL category, but were counted separately under the SD and ELL categories. SD includes students identified as having either an Individualized Education Program or protection under Section 504 of the Rehabilitation Act of 1973. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 and 2011 Science Assessments.

Table A-10. Percentage of eighth-grade public and nonpublic school students identified as students with disabilities (SD) and/or English language learners (ELL) excluded and assessed in NAEP science, as a percentage of identified SD and/or ELL students, by grade and SD/ELL category: 2011

SD/ELL category	Percentage of identified SD and/or ELL students			
	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations
SD and/or ELL	9	91	27	64
SD	12	88	13	75
ELL	6	94	54	40

NOTE: Students identified as both SD and ELL were counted only once under the combined SD and/or ELL category, but were counted separately under the SD and ELL categories. SD includes students identified as having either an Individualized Education Program or protection under Section 504 of the Rehabilitation Act of 1973. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Science Assessment.

State Exclusion Rates (public school students only): The state percentages of eighth-graders identified as SD and/or ELL in 2011 ranged from 8 to 23 percent, and exclusion rates ranged from 1 to 3 percent (table A-11).

Rates by state are reported separately for SD and ELL students at grade 8 in tables A-12 and A-13. Rates are also reported as the percentage of SD and/or ELL students identified in each state in table A-14.

Table A-11. Percentage of eighth-grade public school students with disabilities and/or English language learners identified, excluded, and assessed in NAEP science, as a percentage of all students, by state/jurisdiction: 2009 and 2011

State/jurisdiction	2009					2011				
	Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations	Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations
Nation (public)	18	2	16	5	10	18	2	16	5	11
Alabama	11	1	10	7	3	12	1	11	7	4
Alaska	—	—	—	—	—	21	1	20	4	16
Arizona	16	2	14	5	9	12	1	11	2	9
Arkansas	16	1	14	3	11	16	1	15	3	12
California	25	2	24	18	6	23	2	22	14	8
Colorado	17	1	15	5	11	16	1	15	5	10
Connecticut	16	2	14	3	11	16	1	15	2	13
Delaware	17	1	16	2	14	16	2	14	2	12
Florida	19	2	17	1	16	19	1	17	1	16
Georgia	13	1	12	2	10	12	2	10	2	8
Hawaii	18	2	17	6	10	20	2	18	7	11
Idaho	12	1	11	4	7	12	1	10	4	7
Illinois	16	1	15	3	12	17	1	16	3	12
Indiana	16	2	14	3	12	17	1	16	3	13
Iowa	16	1	15	2	12	17	1	16	2	14
Kansas	—	—	—	—	—	18	1	16	7	9
Kentucky	13	2	10	2	9	13	3	10	2	8
Louisiana	16	1	14	2	12	15	1	14	1	13
Maine	19	2	17	3	14	20	2	18	4	14
Maryland	14	3	12	1	11	14	2	12	1	11
Massachusetts	21	4	17	3	14	22	3	19	3	16
Michigan	15	2	12	3	9	14	3	12	3	8
Minnesota	17	2	15	6	9	17	2	15	7	8
Mississippi	10	1	9	2	7	8	1	7	1	6
Missouri	14	1	12	3	10	14	1	13	3	10
Montana	14	2	12	3	9	13	2	12	3	9
Nebraska	—	—	—	—	—	16	1	15	3	12
Nevada	17	1	16	5	10	18	1	17	6	11
New Hampshire	21	2	19	5	14	20	2	18	5	13
New Jersey	18	2	16	1	14	19	1	18	1	17
New Mexico	21	3	18	8	11	22	2	20	10	10
New York	20	2	18	1	17	20	1	19	#	18
North Carolina	17	2	15	3	13	18	2	16	4	12
North Dakota	16	4	12	3	9	16	3	13	2	10
Ohio	15	2	13	1	12	16	2	14	2	12
Oklahoma	18	3	14	4	10	18	3	15	5	10
Oregon	18	2	16	8	9	18	2	16	6	10
Pennsylvania	19	2	17	2	15	17	1	16	2	15
Rhode Island	21	3	18	4	14	19	1	19	4	14
South Carolina	16	2	14	5	9	15	1	14	5	9
South Dakota	12	1	10	3	7	13	1	11	3	8
Tennessee	12	2	11	1	9	13	1	12	1	10
Texas	17	4	14	7	7	18	2	16	8	8
Utah	14	2	12	4	8	14	2	12	3	9
Vermont	—	—	—	—	—	20	1	18	4	14
Virginia	17	2	15	4	11	18	3	15	5	10
Washington	14	2	12	4	7	16	2	14	5	10
West Virginia	15	2	14	4	10	14	2	12	3	9
Wisconsin	18	2	16	3	13	18	2	16	3	14
Wyoming	15	2	13	3	10	14	1	13	2	11
Other jurisdictions										
District of Columbia	—	—	—	—	—	21	1	20	2	18
DoDEA ¹	13	2	11	3	7	14	1	13	3	10

— Not available. Did not participate at state level in 2009.

Rounds to zero.

¹ Department of Defense Education Activity (overseas and domestic schools).

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 and 2011 Science Assessments.

Table A-12. Percentage of eighth-grade public school students with disabilities indentified, excluded, and assessed in NAEP science, as a percentage of all students, by state/jurisdiction: 2009 and 2011

State/jurisdiction	2009					2011				
	Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations	Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations
Nation (public)	13	2	11	2	9	13	2	11	2	9
Alabama	10	1	9	6	3	10	1	9	5	4
Alaska	—	—	—	—	—	13	1	12	1	11
Arizona	12	2	10	2	7	11	1	10	1	8
Arkansas	12	1	11	2	9	11	1	10	1	9
California	9	1	8	3	5	10	2	8	2	6
Colorado	11	1	9	1	8	10	1	9	1	8
Connecticut	13	1	12	2	10	12	1	11	1	10
Delaware	15	1	14	1	13	14	2	13	2	11
Florida	15	1	14	1	12	14	1	13	1	12
Georgia	11	1	10	2	8	10	2	9	2	7
Hawaii	12	1	11	3	8	11	1	10	2	8
Idaho	9	1	8	3	5	8	1	7	2	5
Illinois	14	1	13	2	11	14	1	13	2	11
Indiana	14	2	12	1	10	14	1	13	2	11
Iowa	14	1	13	1	12	15	1	14	1	13
Kansas	—	—	—	—	—	12	1	10	2	8
Kentucky	12	2	9	1	8	12	2	9	1	8
Louisiana	15	1	13	2	12	14	1	13	1	13
Maine	17	2	16	3	13	18	2	17	3	14
Maryland	12	2	10	1	9	11	2	10	1	9
Massachusetts	19	3	15	2	13	19	3	16	1	14
Michigan	13	2	10	2	8	12	2	10	2	8
Minnesota	12	2	11	3	8	13	2	11	4	7
Mississippi	9	1	8	1	7	7	1	7	1	6
Missouri	13	1	12	3	9	13	1	12	2	10
Montana	12	2	10	1	9	12	2	10	2	9
Nebraska	—	—	—	—	—	14	1	13	2	11
Nevada	11	1	10	2	8	10	1	9	2	7
New Hampshire	20	2	18	5	13	18	2	16	4	12
New Jersey	16	2	14	1	13	17	1	16	1	16
New Mexico	13	3	10	3	7	12	2	11	3	8
New York	16	1	15	1	14	16	1	15	#	14
North Carolina	12	1	11	1	10	14	1	12	2	10
North Dakota	15	4	11	3	9	14	3	11	2	9
Ohio	15	2	12	1	11	15	2	13	1	12
Oklahoma	15	3	12	2	10	16	3	13	4	9
Oregon	13	2	11	5	7	13	2	12	3	9
Pennsylvania	17	2	16	2	14	16	1	14	1	13
Rhode Island	18	2	16	4	12	16	#	16	3	13
South Carolina	14	2	12	4	8	11	1	10	2	8
South Dakota	10	1	9	2	7	11	1	10	2	7
Tennessee	12	2	10	1	9	12	1	10	1	9
Texas	12	3	9	3	6	11	2	9	2	7
Utah	10	2	8	2	7	10	2	9	1	8
Vermont	—	—	—	—	—	18	1	17	3	14
Virginia	14	2	12	3	9	13	2	11	3	8
Washington	11	2	9	3	6	12	2	10	2	8
West Virginia	15	2	13	4	10	14	2	12	3	9
Wisconsin	14	2	12	2	10	14	2	12	2	10
Wyoming	14	1	12	3	10	13	1	12	1	11
Other jurisdictions										
District of Columbia	—	—	—	—	—	17	1	16	1	14
DoDEA ¹	8	1	8	1	6	10	1	9	1	8

— Not available. Did not participate at state level in 2009.

Rounds to zero.

¹ Department of Defense Education Activity (overseas and domestic schools).

NOTE: SD includes students identified as having either an Individualized Education Program or protection under Section 504 of the Rehabilitation Act of 1973. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 and 2011 Science Assessments.

Table A-13. Percentage of eighth-grade public school English language learners identified, excluded, and assessed in NAEP science, as a percentage of all students, by state/jurisdiction: 2009 and 2011

State/jurisdiction	2009					2011				
	Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations	Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations
Nation (public)	6	1	5	3	2	6	#	6	3	2
Alabama	1	#	1	1	#	2	#	1	1	#
Alaska	—	—	—	—	—	11	#	10	3	7
Arizona	6	1	6	3	3	2	#	2	#	2
Arkansas	4	#	4	1	3	5	#	5	2	3
California	20	1	19	16	3	17	1	16	13	4
Colorado	7	#	7	3	3	7	#	7	4	3
Connecticut	4	1	3	1	2	4	#	4	1	3
Delaware	2	#	2	#	2	2	#	2	1	1
Florida	5	1	4	#	4	5	#	4	#	4
Georgia	2	#	2	#	1	2	#	2	#	2
Hawaii	7	1	6	3	3	9	1	8	5	3
Idaho	4	#	4	2	2	4	#	4	2	2
Illinois	3	1	3	1	2	4	#	4	2	2
Indiana	3	#	3	1	1	3	#	3	1	2
Iowa	2	#	2	1	1	3	#	3	1	2
Kansas	—	—	—	—	—	7	#	7	5	2
Kentucky	1	#	1	#	1	1	#	1	#	#
Louisiana	1	#	1	#	1	1	#	1	#	1
Maine	2	#	2	1	1	3	#	3	2	1
Maryland	2	#	2	#	2	3	#	2	#	2
Massachusetts	3	1	2	1	1	4	1	3	1	2
Michigan	2	#	2	2	#	3	#	2	1	1
Minnesota	6	1	5	4	1	5	#	5	3	2
Mississippi	1	#	1	#	#	1	#	1	#	1
Missouri	1	#	1	#	#	1	#	1	1	#
Montana	3	#	3	2	1	2	#	2	1	#
Nebraska	—	—	—	—	—	3	#	3	1	1
Nevada	8	#	8	4	4	10	#	10	4	6
New Hampshire	1	#	1	1	1	2	#	2	1	1
New Jersey	3	1	2	#	2	2	#	2	#	2
New Mexico	11	1	10	5	5	12	1	11	7	4
New York	5	1	4	#	4	6	#	5	#	5
North Carolina	5	#	5	2	3	5	#	4	2	2
North Dakota	2	1	1	1	#	2	#	2	#	2
Ohio	1	#	1	#	#	1	#	1	#	1
Oklahoma	3	#	3	2	1	3	#	3	2	1
Oregon	6	#	6	3	3	6	#	6	3	3
Pennsylvania	2	#	2	1	1	2	#	2	#	2
Rhode Island	3	1	2	1	1	3	#	3	1	2
South Carolina	3	#	3	1	2	5	#	5	2	2
South Dakota	1	#	1	1	#	2	#	2	1	1
Tennessee	1	#	1	#	1	2	#	2	#	1
Texas	7	1	6	4	1	9	1	8	7	1
Utah	5	#	4	2	2	5	#	5	2	3
Vermont	—	—	—	—	—	1	#	1	1	#
Virginia	3	#	3	1	2	6	1	5	3	2
Washington	4	#	3	2	2	5	#	5	3	2
West Virginia	1	#	1	#	#	#	#	#	#	#
Wisconsin	4	1	4	1	3	5	#	5	1	4
Wyoming	1	#	1	#	1	2	#	2	1	1
Other jurisdictions										
District of Columbia	—	—	—	—	—	6	#	5	1	4
DoDEA ¹	5	1	4	2	1	5	1	4	2	2

— Not available. Did not participate at state level in 2009.

Rounds to zero.

¹ Department of Defense Education Activity (overseas and domestic schools).

NOTE: Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2009 and 2011 Science Assessments.

Table A-14. Percentage of eighth-grade public school students identified as students with disabilities (SD) and/or English language learners (ELL) excluded and assessed in NAEP science, as a percentage of identified SD and/or ELL students, by state/jurisdiction: 2011

State/jurisdiction	Percentage of identified SD and/or ELL students											
	SD and/or ELL				SD				ELL			
	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations
Nation (public)	10	90	27	63	12	88	13	75	6	94	54	40
Alabama	9	91	56	35	10	90	52	38	‡	‡	‡	‡
Alaska	5	95	18	77	7	93	10	83	3	97	26	71
Arizona	7	93	14	78	8	92	13	79	‡	‡	‡	‡
Arkansas	6	94	21	73	8	92	12	80	2	98	37	61
California	8	92	59	33	18	82	19	63	4	96	73	23
Colorado	6	94	29	65	9	91	9	82	2	98	52	46
Connecticut	8	92	12	79	10	90	11	79	7	93	15	78
Delaware	10	90	14	76	11	89	11	78	‡	‡	‡	‡
Florida	7	93	6	87	6	94	7	87	7	93	4	89
Georgia	13	87	16	71	15	85	17	69	‡	‡	‡	‡
Hawaii	10	90	35	55	9	91	18	73	11	89	54	36
Idaho	12	88	31	57	16	84	21	63	6	94	49	44
Illinois	7	93	20	73	7	93	12	81	6	94	43	51
Indiana	8	92	16	76	9	91	12	79	2	98	35	63
Iowa	6	94	10	84	6	94	6	88	2	98	29	69
Kansas	8	92	40	52	11	89	19	70	2	98	74	24
Kentucky	21	79	15	64	20	80	13	67	‡	‡	‡	‡
Louisiana	8	92	7	85	8	92	5	87	‡	‡	‡	‡
Maine	9	91	20	70	9	91	17	73	‡	‡	‡	‡
Maryland	14	86	8	78	14	86	6	80	‡	‡	‡	‡
Massachusetts	15	85	13	73	15	85	7	78	17	83	35	48
Michigan	19	81	23	58	19	81	17	64	‡	‡	‡	‡
Minnesota	12	88	38	50	14	86	28	58	5	95	62	33
Mississippi	11	89	14	75	11	89	11	78	‡	‡	‡	‡
Missouri	9	91	19	72	9	91	16	75	‡	‡	‡	‡
Montana	11	89	21	68	13	87	14	74	‡	‡	‡	‡
Nebraska	9	91	21	70	9	91	14	77	6	94	53	41
Nevada	7	93	31	62	11	89	17	72	4	96	41	55
New Hampshire	11	89	23	66	11	89	21	68	‡	‡	‡	‡
New Jersey	6	94	4	90	6	94	3	91	‡	‡	‡	‡
New Mexico	8	92	45	47	13	87	22	65	7	93	60	33
New York	7	93	2	91	7	93	2	91	8	92	2	90
North Carolina	9	91	22	69	11	89	14	75	5	95	42	54
North Dakota	20	80	15	64	23	77	13	64	‡	‡	‡	‡
Ohio	13	87	9	77	14	86	7	78	‡	‡	‡	‡
Oklahoma	16	84	29	55	17	83	23	60	‡	‡	‡	‡
Oregon	9	91	33	58	12	88	24	64	4	96	50	46
Pennsylvania	6	94	9	85	7	93	9	85	‡	‡	‡	‡
Rhode Island	3	97	22	75	3	97	19	78	8	92	35	57
South Carolina	8	92	30	62	10	90	20	70	1	99	52	47
South Dakota	10	90	27	64	10	90	22	69	‡	‡	‡	‡
Tennessee	11	89	11	78	12	88	11	77	‡	‡	‡	‡
Texas	13	87	45	42	18	82	16	66	9	91	75	16
Utah	13	87	19	67	16	84	9	74	5	95	37	58
Vermont	7	93	20	72	7	93	18	75	‡	‡	‡	‡
Virginia	15	85	29	56	15	85	21	64	17	83	42	41
Washington	12	88	28	60	14	86	16	70	4	96	56	40
West Virginia	11	89	23	65	12	88	22	66	‡	‡	‡	‡
Wisconsin	10	90	15	74	13	87	12	75	3	97	24	74
Wyoming	9	91	13	78	10	90	8	82	‡	‡	‡	‡
Other jurisdictions												
District of Columbia	7	93	11	82	7	93	7	86	9	91	22	70
DoDEA ¹	9	91	23	68	6	94	10	84	17	83	47	36

‡ Reporting standards not met. Sample size insufficient to permit a reliable estimate.

¹ Department of Defense Education Activity (overseas and domestic schools).

NOTE: Students identified as both SD and ELL were counted only once under the combined SD and/or ELL category, but were counted separately under the SD and ELL categories. SD includes students identified as having either an Individualized Education Program or protection under Section 504 of the Rehabilitation Act of 1973. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Science Assessment.

Data Collection

The NAEP 2011 science assessment was conducted from January to March 2011 by contractors to the U.S. Department of Education. Data collection for NAEP involves a collaborative effort among the participating schools, school districts, states, and NAEP staff. To reduce the burden on the participating schools, NAEP field staff perform most of the work associated with the assessment. The cooperation of the schools involves enlisting a school staff member to assist in coordinating selected students and providing space to administer the assessments.

Assessment sessions are scripted so that all students are given the same instructions and opportunity to demonstrate what they know and can do. Assessment administrators conduct the sessions under the supervision of their team's assessment coordinator. Training of assessment administrators focuses on their responsibilities in the classroom and on reading the scripts verbatim to administer the sessions in a uniform manner.

NAEP procedures guarantee the anonymity of participants. The names of students are never removed from the schools. The results of NAEP are reported on the national level and by region of the country, state, and for some urban districts—not by school or individual student.

Scoring

Four types of cognitive items were scored for the NAEP science assessment. Responses to multiple-choice questions were scored by high-speed scanners during student booklet processing. Dichotomous constructed-response (correct and incorrect), short constructed-response (correct, partial, and incorrect) and the extended constructed-response questions (those with four or five valid score points) were scored by trained personnel using high-definition images of student responses also captured during processing.

Scoring a large number of short and extended constructed responses with a high level of accuracy and reliability within a limited time frame is essential to the success of NAEP. To ensure reliable, efficient scoring, NAEP does the following:

- develops focused, explicit scoring guides for each item that match the criteria delineated in the assessment frameworks;
- pilot tests all items and adjusts the scoring guides (if necessary) to reflect actual student responses;
- recruits qualified and experienced scorers, trains them, and verifies their ability to score their assigned questions through practice assignments, and in certain cases, qualifying tests;
- employs an image-processing and scoring system that routes images of student responses directly to the scorers so they can focus on scoring rather than paper routing;
- monitors scorer consistency through a second scoring. This procedure randomly selects 5 percent of state samples and 25 percent of the national sample to score twice by different scorers;
- assesses the quality of scorer decision-making through constant monitoring by NAEP assessment experts; and
- documents all training, scoring, and quality control procedures in the technical reports.

For the 2011 science assessment, about one and a half million individual student responses were scored (including second scoring to monitor within-year interrater reliability). There are approximately 3/4 of the 2011 science items that had 90 percent or higher exact agreement between raters of the same student responses. Note that for scoring purposes, each individual part of a multipart item or the bilingual versions of a regular item that is given to the bilingual accommodated students were scored as separate items.

Data Analysis and Scaling

The goal of the analysis of NAEP data is to summarize the performance of groups of students. Initial analysis activities verify the accuracy of the data and data files used in the analysis and provide the first indication of aspects of the data and analysis that require special consideration and attention. The first step is to determine the percentages of students who gave various responses to each cognitive item. Next, the properties of the items are further examined using classical test theory measures of item difficulty and item discrimination. Some of these activities are conducted without student weights or with preliminary student weights, but final student weights are used whenever possible.

After the initial activities are completed, Item Response Theory (IRT) models are used to describe the relationships between the item responses provided by students and the underlying scale. The primary purpose of IRT scaling is to provide a common scale on which performance can be compared, even when students receive different blocks of items. Item parameters that are used in the models are estimated from student response data for each item. Different IRT models with different types of item parameters are used to describe multiple-choice items, dichotomous constructed-response items, and polytomous constructed-response items.

Because the NAEP matrix design gives each student a small proportion of the pool of assessment items, the assessment cannot provide reliable information about individual student performance. Traditional test scores for individual students, even those based on IRT, would result in misleading estimates of population characteristics, such as student group means and percentages of students at or above a certain scale-score level. However, it is NAEP's goal to estimate these population characteristics. NAEP's objectives can be achieved with methodologies that produce estimates of the population-level parameters using marginal estimation techniques for latent variables. Under the assumptions of the analysis models, these population estimates will be consistent in the sense that the estimates approach the population values as the sample size increases.

Prior to 2009, the overall science scale for each grade was a composite scale as a weighted average of subscales estimated for each of the science content areas. Starting in 2009, the overall science scale for each grade is estimated as a single scale. IRT and the NAEP marginal estimation methodology are used to estimate the overall score scale. The overall scale for each grade ranges from 0 to 300, and summarizes student performance across all three science content areas (Physical Science, Life Science, and Earth and Space Sciences) and across all three types of questions in the assessment (multiple choice, short constructed response, and extended constructed response). Summary statistics of the scale scores are estimated, and statistical tests are used to make inferences about the comparisons of results for different groups of students. Finally, NAEP scale score distributions are described via achievement levels and/or item mapping procedures. Additionally, score scales are estimated for each of the three science content areas (Physical Science, Life Science, and Earth and Space Sciences). These subscale scores are also reported on a 0 to 300 scale. For more information about NAEP analysis, IRT, and scaling see <http://nces.ed.gov/nationsreportcard/tdw/analysis/>.

Variance Estimation

The averages and percentages in this report are estimates based on samples of students rather than on entire populations. Moreover, the collection of questions used at each grade level is only a sample of the many questions that could have been asked to assess the skills and abilities described in the NAEP framework, and each assessed student takes only a subset of the entire collection of questions. Therefore, the results are subject to a measure of uncertainty, reflected in the standard error of the estimates—a range of up to a few points above or below the score or percentage—which takes into account potential score fluctuation due to both sampling error and measurement error.

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 errors. 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 science content area, the estimated scale score for any single student would be imprecise. In this case, NAEP's marginal estimation methodology is used to describe the performance of groups of students without requiring precise estimates of individual student performance. The estimate of the variance of the students' scale score distributions (which reflect the imprecision due to lack of measurement accuracy) is computed. This component of variability is then included in the standard errors of NAEP scale scores.

Drawing Inferences from the NAEP Results

Drawing correct inferences from NAEP assessment results depends on the use of appropriate statistical procedures for comparing assessment results for population groups of interest and following guidelines to ensure the validity of the inferences. Comparisons of different groups of students with respect to scores or percentages of a certain attribute are of primary interest to users of NAEP results. The user is cautioned to rely on the results of statistical tests, rather than on the apparent magnitude of the difference between two estimates when determining whether differences are likely to represent actual differences among the groups in the population.

***t* Test Comparison:** By convention, references to differences in NAEP reports indicate that scores or percentages from two groups are different (e.g., one group performed higher or lower than another group) only when the difference in the point estimates for the groups being compared is statistically significant at an approximate level of .05.

Since 1998, *t* tests have been used for most NAEP comparisons. These tests are more appropriate than *z* tests (based on normal distribution approximations) when the statistics that are being compared are from distributions with proportionally larger extremes (i.e., thicker tails) than the normal distribution. One aspect of the use of *t* tests that contributes to the difficulty in their use for large-scale surveys is the determination of the appropriate degrees of freedom for the *t* distribution of interest.

Multiple Comparison Procedures: The *t* test used by NAEP 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, in some sections of a report, many different groups may be compared (i.e., multiple sets of confidence intervals are being analyzed). In sets of confidence intervals, statistical theory indicates that 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., .05), adjustments—called multiple comparison procedures—must be made to the methods.

To ensure that comparisons made using NAEP data are as accurate as possible, error rates are controlled when multiple comparisons are made. When making a number of comparisons in a single analysis, such as analyzing White student performance versus the performance of Black, Hispanic, Asian/Pacific Islander, and American Indian/Alaska Native students, the probability of finding significant differences by chance, for at least one comparison, increases with the family size or number of comparisons. There are several ways to take into account how many related comparisons are being made. In NAEP, the Benjamini-Hochberg False Discovery Rate (FDR) procedure is used to control for this.

Unlike other multiple comparison procedures (e.g., the Bonferroni procedure) that control the familywise error rate (i.e., the probability of making even one false rejection in the set of comparisons), the FDR procedure controls the expected proportion of falsely rejected hypotheses. Familywise procedures are considered conservative for large families of comparisons; therefore the FDR procedure is more suitable for multiple comparisons in NAEP than other procedures. There are two exceptions where the FDR is not applied: when comparing multiple years and when comparing multiple jurisdictions to the nation.

NAEP Reporting Groups

In addition to overall results assessed, NAEP results are reported for certain student groups provided there are sufficient numbers of students and adequate school representation. Results for some student groups may not be available for certain years, grades, or jurisdictions.

Race/Ethnicity: The school-recorded race/ethnicity variable records the race/ethnicity of each student as reported by the student's school. When the school-recorded information is missing, student-reported data derived from the student background questions are used. For 2011, the mutually exclusive racial/ethnic categories are White, Black, Hispanic, Asian, American Indian/Alaska Native, Native Hawaiian or Other Pacific Islander, and Two or more races. Black includes African American and Hispanic includes Latino. Race categories exclude Hispanic origin unless specified.

Gender: The gender of the student assessed is taken from school records.

Eligibility for the National School Lunch Program: The school lunch variable is based on available school records. Students are classified as either currently eligible or not currently eligible for the national lunch component of the Department of Agriculture's National School Lunch Program. The classification refers only to the school year when the assessment was administered and is not based on eligibility in previous years. If school records are not available, the student is classified as "Information not available." If the school did not participate in the program, all students in that school were classified as "Information not available." Eligibility for the program is determined by students' family income in relation to the federally established poverty level. Free lunch qualification is set at 130 percent of the poverty level or below, and reduced-price lunch qualification is set at between 130 and 185 percent of the poverty level. (For the period July 1, 2010 through June 30, 2011, for a family of four, 130 percent of the poverty level was \$28,665, and 185 percent was \$40,793.) Additional information on eligibility may be found at the U.S. Department of Agriculture website at <http://www.fns.usda.gov/cnd/lunch/>.

Type of Location: Results for four mutually exclusive categories of school location are also reported: city, suburb, town, and rural. The categories are based on standard definitions established by the Federal Office of Management and Budget using population and geographic information from the U.S. Census Bureau. Schools are assigned to these categories in the NCES Common Core of Data based on their physical address.

Parental Education: Eighth-graders assessed in 2011 were asked the following two questions, the responses to which were combined to derive the parental education variable:

How far in school did your mother go?

- She did not finish high school.
- She graduated from high school.
- She had some education after high school.
- She graduated from college.
- I don't know.

How far in school did your father go?

- He did not finish high school.
- He graduated from high school.
- He had some education after high school.
- He graduated from college.
- I don't know.

The information was combined into one parental-education reporting variable in the following way:

- If a student indicated the extent of education for only one parent, that level was included in the data. If a student indicated the extent of education for both parents, the higher of the two levels was included in the data.

- If a student responded "I don't know" for both parents, or responded "I don't know" for one parent and did not respond for the other, the parental education level was classified as "I don't know."
- If the student did not respond for either parent, the student was recorded as having provided no response.

Region of the Country: Prior to 2003, NAEP results were reported for four NAEP-defined regions of the nation: Northeast, Southeast, Central, and West. To align NAEP with other federal data collections, NAEP analysis and reports have used the U.S. Census Bureau's definition of "region" beginning in 2003. The four regions defined by the U.S. Census Bureau are Northeast, South, Midwest, and West. Therefore, trend data by region are not provided for assessment years prior to 2003.

Figure A-1 shows how states are subdivided into these census regions. All 50 states and the District of Columbia are listed. Other jurisdictions, including the Department of Defense Education Activity schools, are not assigned to any region.

Figure A-1. States within regions of the country defined by the U.S. Census Bureau

Northeast	South	Midwest	West
Connecticut	Alabama	Illinois	Alaska
Maine	Arkansas	Indiana	Arizona
Massachusetts	Delaware	Iowa	California
New Hampshire	District of Columbia	Kansas	Colorado
New Jersey	Florida	Michigan	Hawaii
New York	Georgia	Minnesota	Idaho
Pennsylvania	Kentucky	Missouri	Montana
Rhode Island	Louisiana	Nebraska	Nevada
Vermont	Maryland	North Dakota	New Mexico
	Mississippi	Ohio	Oregon
	North Carolina	South Dakota	Utah
	Oklahoma	Wisconsin	Washington
	South Carolina		Wyoming
	Tennessee		
	Texas		
	Virginia		
	West Virginia		

Source: U.S. Department of Commerce Economics and Statistics Administration, U.S. Census Bureau.

Caution in Interpretations

As previously stated, the NAEP science scale makes it possible to examine relationships between students' performance and various background factors that NAEP measures. However, the relationship between achievement and another variable does not reveal its underlying cause, which may be influenced by a number of other variables. Similarly, the assessments do not reflect the influence of unmeasured variables. The results are most useful when considered in combination with other knowledge about the student population and the educational system, such as trends in instruction, changes in the school-age population, and societal demands and expectations.

Caution in interpretation is also warranted for some small population group estimates. At times in this report, smaller population groups show very large increases or decreases across years in average scores; however, it is necessary to interpret such score changes with extreme caution. The effects of exclusion-rate changes for small student groups may be more marked for small groups than they are for the whole population. In addition, standard errors are often quite large around the score estimates for small groups, which in turn means the standard error around the gain is also large.