

Appendix

Overview

This appendix provides an overview of the NAEP 2007 writing assessment's primary components—framework, development, administration, scoring, and analysis. The information provided about the state and national assessments covers grades 8 and 12 (grade 4 was not assessed in 2007, and grade 12 was assessed only with a national, not a state sample), as well as NAEP's Trial Urban District Assessment (TUDA). References marked within the text appear as end notes on the last page of this appendix.

The NAEP Writing Assessment

The National Assessment Governing Board, created by Congress in 1988, is responsible for formulating policy for NAEP. The Governing Board is specifically charged with developing assessment objectives and test specifications. The NAEP 2007 writing assessment is based on the 1998 writing assessment framework that also guided the 2002 assessment. The framework describes the basis for the Writing Assessment, the types of writing assessed, and the methods for scoring student responses. This framework was developed during 1989–90 in preparation for the 1992 NAEP Writing Assessment. For the 1998, 2002, and 2007 assessments, the framework was augmented by a set of Writing Assessment and Exercise Specifications developed during 1995–96. A new framework is under development for the 2011 writing assessment.

Based on the expert opinions of writing educators and researchers, the framework's purpose is to provide a definition of writing upon which the NAEP writing assessment can be based. The framework development process involved the critical input of hundreds of individuals across the country, including representatives of national education organizations, teachers, parents, policymakers, business leaders, and the interested general public. The process was managed by the Center for Research on Evaluation, Standards, and Student Testing (CRESST) for the National Assessment Governing Board, and the test question specifications were developed under contract by American College Testing (ACT) for the Governing Board.

The assessment is designed around the following six overarching objectives:

- Students should write for a variety of purposes (narrative, informative, and persuasive).
- Students should write on a variety of tasks (letters, essays, stories, reports) and for many different audiences (peers, school or government officials, business representatives).
- Students should write from a variety of stimulus materials and within various time constraints.
- Students should generate, draft, revise, and edit ideas and forms of expression in their writing.
- Students should display effective choices in the organization of their writing. They should include detail to illustrate and elaborate their ideas and use appropriate conventions of written English.
- Students should value writing as a communicative activity.

Figure A-1 gives examples of various writing tasks similar to those included in the assessment at grades 8 and 12. Included in the figure are descriptions of sample tasks that illustrate how each purpose for writing is assessed.

Figure A-1

Illustrative examples of NAEP writing tasks at grades 8 and 12, by purpose for writing

Purposes for writing	Grade 8	Grade 12
Narrative	Provide visual stimuli. Ask students to write an article for a sports magazine telling the story of a time when they participated in a hobby or skill they enjoyed.	Provide an appropriate quotation. Ask students to write a letter to a friend telling the story of a time in their lives when they had to make an important decision.
Informative	Provide a series of brief journal entries from another historical time. Ask students to explain what is revealed about the person who wrote the entries.	Provide quotations from a political campaign. Ask students to choose one and in an essay inform their social studies teacher what it means in the context of the campaign.
Persuasive	Provide brief reviews, as models, of a film, TV program, or book. Ask students to write a review for the school newspaper that will convince other students to watch a favorite film or TV program or read a favorite book.	Provide a quotation on education in the United States. Ask students to write a letter to the editor of their local newspaper taking a position on some aspect of education and support it from their own experiences.

SOURCE: National Assessment Governing Board. *Writing Framework and Specifications for the 1998 National Assessment of Education Progress*. Washington, DC: Author.

The framework specifies the percentage of the writing tasks in the assessment that should be devoted to each of the three writing purposes—narrative, informative, and persuasive. The actual distributions of writing tasks in the 2007 assessment are listed in table A-1.

Table A-1

Target percentage of assessment time in NAEP writing and actual number of NAEP writing tasks, by grade and purpose of writing: 2007

Purposes of writing	Grade 8		Grade 12	
	Target percentage of assessment time	Number of tasks	Target percentage of assessment time	Number of tasks
Narrative	33	6	25	4
Informative	33	6	35	6
Persuasive	33	5	40	7

NOTE: NAEP writing was not assessed at grade 4 in 2007. 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), 2007 Writing Assessment.

The Assessment Design

Each student who participated in the writing assessment received a booklet containing two 25-minute writing tasks. All student responses to the writing tasks were rated according to a six-level scoring guide. In addition, the test booklets contained general background questions and writing-specific background questions.

The assessment design allowed for maximum coverage of the writing domain at each grade, while minimizing the time burden for any one student. This was accomplished through the use of matrix sampling of tasks, in which each student was given only 2 of the 17 tasks at each grade level. Representative samples of students responded to each task, so that the aggregate results across the entire assessment allow broad reporting of writing abilities for the targeted population.

In addition to matrix sampling, the assessment design utilized a procedure for distributing blocks across booklets that controlled for position and context effects. Students received different blocks of tasks in their booklets according to a procedure called "partially balanced incomplete block (PBIB) spiraling." The procedure assigned blocks of questions in a manner that balanced the positioning of blocks across booklets and balanced the pairing of blocks within booklets according to purposes for writing. Blocks were balanced within each purpose for writing and were partially balanced across purposes for writing. (The spiraling aspect of this procedure cycles the booklets for administration so that, typically, only a few students in any assessment session receive the same booklets.)

In addition to the student assessment booklets, three other instruments provided data relating to the assessment: a teacher questionnaire, a school questionnaire, and questionnaires about students with disabilities (SD) and/or English language learners (ELL). The teacher questionnaire was administered to teachers of eighth-grade students participating in the assessment. The questionnaire focused on the teacher's general background and experience, the teacher's background related to writing, and the type of classroom organization. The school questionnaire was given to the principal or other administrator in each participating school. The questions asked about school policies, programs, facilities, and the demographic composition and background of the students and teachers at the school.

The SD and ELL questionnaires were completed by a school staff member knowledgeable about those students selected to participate in the assessment who were identified as having an Individualized Education Program (IEP) or equivalent plan (for reasons other than being gifted or talented) or as being an English language learner. An SD or ELL questionnaire was completed for each identified student in the NAEP sample. Each SD or ELL questionnaire asked about the student (for example, type of disability or language spoken other than English) and the special instructional programs (i.e., proportion of time spent in mainstream/general education classes or specially designed instruction) in which the student participated.

NAEP Samples

National Sample

The national results presented in this report are based on nationally representative probability samples of eighth- and twelfth-grade students. The samples were chosen using a stratified two-stage design that involved sampling students from selected schools.

At grade 8 the national sample consisted of the combined sample of public school students assessed in each state that participated in the NAEP state assessment program in writing (including Department of Defense schools in the U.S. and overseas), a representative sample of public schools from the remaining states (including the District of Columbia), and additional nonpublic school samples, covering private schools and Bureau of Indian Affairs schools. The approach of integrating the national and state samples has been used in NAEP since 2002. Prior to 2002, separate samples were drawn for the NAEP national and state assessments. For grade 12, national samples of public and nonpublic schools were selected from across the country, by including a sample of schools and students from each state and the District of Columbia.

For 2007, the sampling frame for public schools was the Common Core of Data (CCD) file corresponding to the 2004–05 school year. The CCD file provided the frame for all regular public, state-operated public, Bureau of Indian Education, and Department of Defense schools that were open during the 2004–05 school year. The sampling frame for private schools was developed from the 2003–04 Private School Survey (PSS), which was carried out by the U.S. Census Bureau for the National Center for Education Statistics (NCES). The PSS is a biennial mail survey of all private schools in the 50 states and the District of Columbia. Supplemental samples of newly-opened public and Catholic schools, drawn from lists other than those of the CCD and PSS, were also selected so as to ensure maximum coverage of the target population.

Each selected school that participated in the assessment and each student assessed represents a portion of the population of interest. Sampling weights are needed to make valid inferences from the student samples to the respective populations from which they were drawn. While each state's NAEP sample provides a representative sample for that state, no state's sample is exactly proportionate to its share of the nation's student population as a whole. Sampling weights compensate for the disproportionate state samples and do the same for the sample of students attending nonpublic schools. Sampling weights also account for lower sampling rates for very small schools and are used to adjust for school and student nonresponse.

For the 2007 national writing assessment, as for the 1998 and 2002 assessments, accommodations for students with disabilities (SD) and English language learners (ELL) were permitted for the entire sample of students. In 2007, accommodations were offered when a student had an Individualized Education Program (IEP) indicating the need for accommodations because of a disability, or was protected under Section 504 of the Rehabilitation Act of 1973 because of a disability, or was identified as being an English language learner, or was normally offered accommodations in other assessment situations.¹ All other students were asked to participate in the assessment under standard conditions.

Table A-2

Sample sizes and target populations in NAEP writing, by grade and type of school: 2007

Type of school	Grade 8		Grade 12	
	Sample size	Target population	Sample size	Target population
Nation	145,200	3,903,000	28,900	3,093,000
Public	140,300	3,554,000	23,000	2,806,000
Nonpublic	4,800	349,000	5,900	287,000

NOTE: The sample size is rounded to the nearest hundred. The target population is rounded to the nearest thousand. Nonpublic school includes private, Bureau of Indian Education, and Department of Defense schools. 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), 2007 Writing Assessment.

Table A-2 shows the sample sizes and target populations for the nation, including both public and nonpublic students. All state reports are based on public school students only. The sample sizes and target populations for the 2007 writing assessment are listed for the nation (public) and states in table A-3. In 2005 and 2007, Department of Defense Education Activity (DoDEA) schools are reported as a single jurisdiction; in past years, domestic (Department of Defense Domestic Dependent Elementary and Secondary Schools or DDESS) and overseas (Department of Defense Dependents Schools or DoDDS) schools were considered separate jurisdictions.

Table A-3**Sample sizes and target populations of eighth-grade public school students in NAEP writing, by state: 2007**

State/jurisdiction	Sample size	Target population
Nation (public)	140,300	3,554,000
Alabama	2,800	57,000
Arizona	2,700	72,000
Arkansas	2,400	34,000
California	8,300	471,000
Colorado	2,700	57,000
Connecticut	2,500	41,000
Delaware	2,800	10,000
Florida	4,000	192,000
Georgia	3,600	120,000
Hawaii	2,700	13,000
Idaho	2,900	21,000
Illinois	4,000	152,000
Indiana	2,700	80,000
Iowa	2,800	37,000
Kansas	2,800	33,000
Kentucky	2,700	46,000
Louisiana	2,400	48,000
Maine	2,600	15,000
Massachusetts	3,700	69,000
Michigan	2,600	121,000
Minnesota	2,900	61,000
Mississippi	2,600	36,000
Missouri	2,800	71,000
Montana	2,600	11,000
Nevada	2,600	28,000
New Hampshire	2,800	16,000
New Jersey	2,800	104,000
New Mexico	2,800	26,000
New York	3,800	206,000
North Carolina	4,200	104,000
North Dakota	2,300	8,000
Ohio	3,700	136,000
Oklahoma	2,600	43,000
Pennsylvania	2,800	139,000
Rhode Island	2,600	12,000
South Carolina	2,700	50,000
Tennessee	2,800	74,000
Texas	7,300	299,000
Utah	2,700	35,000
Vermont	2,000	7,000
Virginia	2,800	90,000
Washington	3,000	77,000
West Virginia	2,900	22,000
Wisconsin	2,700	62,000
Wyoming	1,900	7,000
Other jurisdictions		
BIE ¹	100	3,000
DoDEA ²	1,600	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. Alaska, the District of Columbia, Maryland, Nebraska, Oregon, and South Dakota did not participate in the 2007 NAEP writing assessment, but they were included in the nationally representative sample. Data for BIE and DoDEA schools are not counted in the national (public) totals. 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), 2007 Writing Assessment.

In the 2007 assessment, as in the 2002, 2003, and 2005 NAEP assessments, a number of large urban school districts participated on a voluntary basis in a Trial Urban District Assessment (TUDA), and larger than normal NAEP samples were drawn in these districts to permit reliable reporting of student group performance. Reports from these Trial Urban District Assessments for 2002 in reading and writing, and 2003, 2005, and 2007 in reading and mathematics, are available on the NAEP website at <http://nces.ed.gov/nationsreportcard/>; a report for writing in 2007 is forthcoming. The sample sizes and target populations for the districts participating in TUDA are given in table A-4.

Table A-4

Student sample sizes and target populations of eighth-grade public school students for Trial Urban District Assessment in writing, by urban district: 2007

District	Sample size	Target population
Atlanta	900	3,000
Austin	1,500	5,000
Boston	1,200	4,000
Charlotte	1,400	9,000
Chicago	1,800	26,000
Cleveland	1,200	4,000
Houston	2,100	13,000
Los Angeles	2,000	52,000
New York City	2,000	69,000
San Diego	1,400	10,000

NOTE: The sample size is rounded to the nearest hundred. The target population is rounded to the nearest thousand.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Trial Urban District Writing Assessment.

State Samples

The results of the 2007 state assessment in writing provided in this report are based on state-level samples of eighth-grade public school students. The samples were selected using a two-stage sample design that first selected schools within each state or other jurisdiction and then selected students within schools. The samples were weighted to allow valid inferences about the populations of interest. Participation rates for the states and other jurisdictions were calculated the same way that rates were computed for the nation. Table A-5 displays weighted school and student participation rates for the state samples at grade 8. The student-weighted school participation rates are calculated based on the estimated number of students represented by either the initially selected schools that participated in the assessment (before substitution) or the participating schools, whether originally selected or selected as a substitute for a school that chose not to participate (after substitution). The school-weighted school participation rates are calculated based on the estimated number of schools either before or after substitutions.

Table A-5

Public school and student participation rates in NAEP writing at grade 8, by state: 2007

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	6,310	92	135,100
Alabama	100	100	120	93	2,700
Arizona	100	100	130	91	2,600
Arkansas	100	100	120	92	2,400
California	100	100	310	92	8,100
Colorado	96	98	120	92	2,600
Connecticut	97	97	100	91	2,500
Delaware	100	100	50	92	2,600
Florida	100	100	160	91	3,900
Georgia	100	100	120	93	3,500
Hawaii	100	100	70	92	2,700
Idaho	99	99	110	94	2,800
Illinois	100	100	200	93	3,900
Indiana	100	100	110	92	2,600
Iowa	100	100	140	93	2,800
Kansas	100	100	150	93	2,700
Kentucky	100	100	110	93	2,500
Louisiana	100	100	110	92	2,300
Maine	96	98	130	92	2,500
Massachusetts	100	100	140	91	3,400
Michigan	100	100	120	92	2,500
Minnesota	98	99	140	91	2,800
Mississippi	100	100	110	93	2,600
Missouri	100	100	130	94	2,800
Montana	100	98	170	92	2,500
Nevada	100	100	80	87	2,500
New Hampshire	98	98	90	91	2,700
New Jersey	98	97	110	92	2,700
New Mexico	100	100	110	90	2,600
New York	100	100	160	90	3,600
North Carolina	100	100	150	92	4,000
North Dakota	99	98	180	94	2,100
Ohio	100	100	190	92	3,500
Oklahoma	100	100	150	92	2,500
Pennsylvania	100	100	110	92	2,700
Rhode Island	100	100	60	91	2,600
South Carolina	100	100	110	93	2,600
Tennessee	100	100	120	93	2,700
Texas	100	100	220	93	6,800
Utah	100	100	110	91	2,600
Vermont	100	100	120	94	2,000
Virginia	100	100	110	93	2,600
Washington	100	100	130	91	2,800
West Virginia	100	100	120	93	2,800
Wisconsin	98	98	130	92	2,600
Wyoming	100	100	80	92	1,800
Other jurisdiction					
DoDEA ¹	100	98	70	92	1,600

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

NOTE: The numbers of schools are rounded to the nearest ten, and the numbers of students are rounded to the nearest hundred. The percentages for student-weighted school participation and school-weighted school participation have different denominators; see accompanying text for definitions. Substitutions of reserve schools for initially sampled schools were not needed in 2007 because school participation rates were high. Alaska, the District of Columbia, Maryland, Nebraska, Oregon, and South Dakota did not participate in the 2007 NAEP writing assessment, but they were included in the nationally representative sample. 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), 2007 Writing Assessment.

District Samples

Results from the 2007 writing assessment are also reported for district-level samples of eighth-grade students in the large urban school districts that participated in the TUDA—Atlanta City, Austin, Boston School District, Charlotte-Mecklenburg Schools, City of Chicago School District 299, Cleveland Municipal School District, Houston Independent School District, Los Angeles Unified, New York City Public Schools, and San Diego City Unified. The District of Columbia, which is regularly included in NAEP assessments as a jurisdiction, could not participate in the writing assessment because it did not have enough students to participate simultaneously in the reading, mathematics, and writing assessments in 2007. The sample of students in the urban school districts represents an augmentation of the sample of students who would usually be selected as part of the state samples. These samples allow reliable reporting of student groups within these districts. Furthermore, all students at more local geographic sampling levels are assumed to be part of broader samples. For example, Houston is one of the urban districts included in the TUDA. Data from students tested in the Houston sample were used to report results for Houston, but also contributed to the Texas and national estimates. Participation rates for the urban district samples are presented in table A-6.

Table A-6

Public school and student participation rates for Trial Urban District Assessment in writing at grade 8, by urban district: 2007

District	School participation		Student participation	
	Student-weighted percent	Number of schools participating	Student-weighted percent	Number of students assessed
Atlanta	100	20	91	900
Austin	100	20	92	1,400
Boston	100	30	91	1,100
Charlotte	100	30	90	1,300
Chicago	100	100	95	1,700
Cleveland	100	80	87	1,100
Houston	100	50	92	1,900
Los Angeles	100	70	91	2,000
New York City	100	80	88	1,900
San Diego	100	30	93	1,400

NOTE: The numbers of schools are rounded to the nearest ten, and the numbers of students are rounded to the nearest hundred. Substitutions of reserve schools for initially sampled schools were not needed in 2007 because school participation rates were high. The percentages for school-weighted and student-weighted school participation are both at 100 percent for the participating districts in 2007.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2007 Trial Urban District Writing Assessment.

Standards for State Sample Participation and Reporting of Results

In carrying out the 2007 state assessment program, the National Center for Education Statistics (NCES) established participation rate standards that states and other jurisdictions were required to meet for their results to be reported. NAEP state assessment results are based only on public schools. Participation rates before substitution needed to be at least 85 percent for schools and at least 85 percent for students. In the 2007 writing assessment at grade 8, all participating jurisdictions met NCES participation rate standards. Further information on the NCES guidelines used to report results in the state assessments, and the guidelines for notations when there was some risk of nonresponse bias in the reported results prior to the 2003 assessments, can be found in the NAEP 2002 writing report card (NCES 2003–529; see appendix A, "Standards for Sample Participation and Reporting of Results").

Students With Disabilities (SD) and/or English Language Learners (ELL)

It is important to assess all selected students from the target population. Therefore, every effort is made to ensure that all selected students who are capable of participating in the assessment are assessed. Some students sampled for participation in NAEP can be excluded from the sample according to carefully defined criteria. These criteria were revised in 1996 to communicate more clearly a presumption of inclusion except under special circumstances. According to these criteria, students who had an Individualized Education Program (IEP) or were protected under Section 504 of the Rehabilitation Act of 1973 were to be included in the NAEP assessment.

In the 2007 assessment, procedures for including SD and ELL students were further refined. School staff made the decisions about whether to include an SD or ELL student in a NAEP assessment and which testing accommodations, if any, they should receive. All ELL students are assessed in NAEP the same way they are in their state assessments. If an ELL student takes a simplified English or native language academic assessment, NAEP staff work with the school to determine if the student could take NAEP assessments with any of the allowable accommodations. The NAEP program furnishes tools to assist school personnel in making those decisions.

A sampling procedure was used to select students at each grade being tested. Students were selected on a random basis, without regard to SD or ELL status. Once the students were selected, the schools identified those who had SD or ELL status. School staff who were familiar with these students were asked a series of questions to help them decide whether each student should participate in the assessment and whether the student needs accommodations.

Inclusion in NAEP of an SD or ELL student is encouraged (a) if that student participated in the regular state academic assessment in the subject being tested, and (b) if that student can participate in NAEP with the accommodations NAEP allows. Even if the student did not participate in the regular state assessment, or if he/she needs accommodations NAEP does not allow, school staff are asked whether that student could participate in NAEP with the allowable accommodations. (For example, extending testing over several days is not allowed for NAEP because NAEP administrators are in each school for only one day.)

Participation of SD/ELL Students in the NAEP Samples

Testing all sampled students is the best way for NAEP to ensure that the statistics generated by the assessment are as representative as possible of the performance of the entire national population and the populations of participating jurisdictions. However, all groups of students include certain proportions that cannot be tested in large-scale assessments (such as students who have profound mental disabilities) or who can only be tested through the use of testing accommodations such as extra time, one-on-one administration, or use of magnifying equipment. Some students with disabilities and some English language learners cannot show on a test what they know and can do unless they are provided with accommodations. When such accommodations are not allowed, students requiring such adjustments are often excluded from large-scale assessments such as NAEP. This phenomenon has become more common since the 1990s, particularly with the passage of the 1997 Individuals with Disabilities Education Act (IDEA), which led schools and states to identify increasing proportions of students as needing accommodations on assessments to best show what they know and can do.² Furthermore, section 504 of the Rehabilitation Act of 1973 requires that, when students with disabilities are tested, schools must provide them with appropriate accommodations so that the test results accurately reflect students' achievement. In addition, as the proportion of ELL students in the population has increased, some states have started offering accommodations such as translations of assessments or the use of bilingual dictionaries as part of the assessments.

Before 1996, no testing under nonstandard conditions was allowed in NAEP, and accommodations were not permitted. At that time, NAEP samples were able to include almost all sampled students in standard assessment sessions. However, as the influence of IDEA became more widespread, the failure to provide accommodations led to increasing levels of exclusion in the assessment. Such increases posed two threats to the program: they threatened the stability of trend lines (because excluding more students in one assessment year than in another might lead to apparent rather than real differences), and they made NAEP samples less than optimally representative of target populations.

A multipart strategy was adopted as a response to this challenge. The program had to move toward allowing the same assessment accommodations that were afforded students in state and district testing programs for NAEP samples to be as inclusive as possible. However, to allow accommodations represents a change in testing conditions that might affect measurement of changes over time. Therefore, beginning with the 1996 national assessments (in mathematics and science) and the 1998 state assessments in reading, and up to 2000, NAEP assessed a series of parallel samples of students. In one set of samples, testing accommodations were not permitted; this allowed NAEP to maintain the measurement of achievement trends. Parallel samples in which accommodations were permitted were also assessed. By having two overlapping samples³ and two sets of related data points, NAEP could meet two core program goals. First, data trends could be maintained. Second, parallel trend lines could be reported during the interim until the program transitioned to a sample with accommodations permitted as its only reporting format. Starting in 2002, NAEP has used only the more inclusive samples, in which assessment accommodations are permitted. In writing, all national and state data from 1998 onward have been conducted with accommodations permitted.

To make it possible to evaluate both the impact of increasing exclusion rates in some jurisdictions and differences between jurisdictions, complete data on exclusion in all years are included in this appendix. Because the exclusion rates may affect trend measurement within a jurisdiction, readers should consider the magnitude of exclusion rate changes when interpreting score changes in jurisdictions. In addition, different rates of exclusion may influence the meaning of state comparisons. Thus, exclusion data should be reviewed in this context as well.

Table A-7 presents the percentages of all public and nonpublic school students who were identified as students with disabilities (SD) or as English language learners (ELL), or both. The table also includes the percentages of all students who were excluded SD and/or ELL and the percentages of all students who were assessed SD and/or ELL for those assessments. The denominator for these percentages includes assessed students plus excluded students; it does not include sampled students who were absent or refused to participate. Tables A-8, A-9, and A-10 show similar information by state for SD/ELL, SD, and ELL, respectively.

Table A-11 presents the identification, exclusion, and accommodation information for the districts that participated in the Trial Urban District Assessment.

In the 2007 national sample, 3 percent of students at grade 8 were excluded from the assessment (see table A-7). Across the various jurisdictions that participated in the 2007 state assessment, the percentage of students excluded ranged from 1 to 7 percent at grade 8 (see table A-8). At the district level, between 2 and 11 percent of students were excluded at grade 8 (see table A-11).

Table A-7

Public and nonpublic school students with disabilities (SD) and English language learners (ELL) identified, excluded, and assessed in NAEP writing, as a percentage of all students, by grade and SD/ELL category: 1998, 2002, and 2007

Grade and SD/ELL category	1998	2002	2007
Grade 8			
SD and/or ELL			
Identified	13	17	17
Excluded	4	4	3
Assessed	9	13	14
Without accommodations	6	8	6
With accommodations	3	5	8
SD			
Identified	10	12	12
Excluded	3	3	3
Assessed	7	9	10
Without accommodations	5	5	2
With accommodations	3	5	7
ELL			
Identified	3	6	6
Excluded	1	1	1
Assessed	2	4	5
Without accommodations	2	4	4
With accommodations	#	1	2
Grade 12			
SD and/or ELL			
Identified	8	11	13
Excluded	2	3	3
Assessed	6	8	10
Without accommodations	5	6	4
With accommodations	1	3	6
SD			
Identified	6	9	10
Excluded	2	3	3
Assessed	4	6	7
Without accommodations	3	4	2
With accommodations	1	3	5
ELL			
Identified	2	3	4
Excluded	#	1	1
Assessed	2	2	3
Without accommodations	2	2	2
With accommodations	#	#	1

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. 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), 1998, 2002, and 2007 Writing Assessments.

Table A-8

Eighth-grade public school students with disabilities and/or English language learners identified, excluded, and assessed in NAEP writing, as a percentage of all students, by state: 1998, 2002, and 2007

State/jurisdiction	1998					2002				
	Identified	Excluded	Assessed	Assessed without accom- modations	Assessed with accom- modations	Identified	Excluded	Assessed	Assessed without accom- modations	Assessed with accom- modations
Nation (public)	14	4	10	7	3	18	4	14	8	5
Alabama	12	6	6	5	1	15	3	12	11	1
Alaska	—	—	—	—	—	—	—	—	—	—
Arizona	17	5	12	10	2	22	5	17	14	3
Arkansas	13	6	7	5	1	17	3	14	9	5
California	23	6	17	15	2	27	3	24	20	3
Colorado	13	4	9	6	3	—	—	—	—	—
Connecticut	15	7	8	5	3	17	4	13	7	6
Delaware	14	3	11	8	3	15	5	11	2	8
Florida	16	5	11	9	2	20	4	16	7	10
Georgia	11	5	7	4	2	13	3	10	5	5
Hawaii	15	4	11	8	3	21	3	18	11	7
Idaho	—	—	—	—	—	14	2	13	8	4
Illinois	12	4	8	6	2	18	3	14	8	7
Indiana	—	—	—	—	—	13	3	10	7	3
Iowa	—	—	—	—	—	—	—	—	—	—
Kansas	—	—	—	—	—	16	3	13	6	7
Kentucky	10	2	7	3	4	11	4	8	4	3
Louisiana	13	5	8	3	5	16	4	12	4	8
Maine	14	5	8	5	3	18	2	16	8	8
Maryland	13	2	11	4	7	16	4	12	9	3
Massachusetts	17	5	12	7	5	20	3	16	7	10
Michigan	—	—	—	—	—	14	5	9	4	4
Minnesota	14	3	11	8	3	17	3	14	9	5
Mississippi	9	5	5	4	1	10	5	5	3	2
Missouri	13	3	10	6	4	16	3	13	4	9
Montana	11	2	9	6	2	13	2	12	7	4
Nebraska	—	—	—	—	—	17	4	12	7	5
Nevada	16	6	10	8	3	21	4	16	12	5
New Hampshire	—	—	—	—	—	—	—	—	—	—
New Jersey	—	—	—	—	—	—	—	—	—	—
New Mexico	23	6	17	14	3	32	5	27	17	10
New York	15	5	9	3	6	20	6	14	5	9
North Carolina	14	4	10	4	6	17	5	12	4	9
North Dakota	—	—	—	—	—	15	1	14	8	6
Ohio	—	—	—	—	—	12	6	7	4	2
Oklahoma	13	9	5	4	1	16	2	14	9	4
Oregon	15	3	12	9	3	18	4	14	11	3
Pennsylvania	—	—	—	—	—	14	2	12	4	8
Rhode Island	17	4	13	10	3	22	3	18	9	10
South Carolina	12	5	7	5	2	15	5	10	6	4
South Dakota	—	—	—	—	—	—	—	—	—	—
Tennessee	13	4	9	8	1	14	3	12	10	2
Texas	19	6	13	10	2	19	7	13	11	2
Utah	10	4	6	5	1	17	3	14	9	4
Vermont	—	—	—	—	—	17	4	14	6	7
Virginia	14	4	9	6	3	18	6	12	5	7
Washington	13	4	9	7	3	15	3	11	6	5
West Virginia	14	5	9	5	3	18	4	14	5	9
Wisconsin	11	4	7	4	3	17	4	13	4	9
Wyoming	9	2	7	5	2	15	2	13	6	7
Other jurisdictions										
District of Columbia	13	6	7	6	1	21	6	15	5	10
DoDEA ¹	8	2	6	4	2	11	2	9	6	4

See notes at end of table.

Table A-8

Eighth-grade public school students with disabilities and/or English language learners identified, excluded, and assessed in NAEP writing, as a percentage of all students, by state: 1998, 2002, and 2007—Continued

State/jurisdiction	2007				
	Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations
Nation (public)	18	3	15	6	9
Alabama	14	2	12	8	3
Alaska	—	—	—	—	—
Arizona	19	3	16	10	6
Arkansas	16	2	14	4	10
California	27	2	25	20	5
Colorado	15	3	12	4	8
Connecticut	14	2	13	3	9
Delaware	16	5	11	3	8
Florida	19	3	17	2	15
Georgia	13	2	11	3	8
Hawaii	19	1	18	8	11
Idaho	14	2	12	6	6
Illinois	17	3	14	3	11
Indiana	16	3	13	3	10
Iowa	16	2	15	3	12
Kansas	17	4	13	4	9
Kentucky	15	6	9	2	6
Louisiana	14	2	12	1	10
Maine	20	4	16	4	12
Maryland	—	—	—	—	—
Massachusetts	22	6	16	3	13
Michigan	16	4	12	3	9
Minnesota	17	2	15	6	9
Mississippi	10	2	9	1	7
Missouri	15	2	12	3	9
Montana	16	2	14	4	10
Nebraska	—	—	—	—	—
Nevada	21	3	18	11	7
New Hampshire	20	3	17	5	12
New Jersey	18	3	15	2	13
New Mexico	27	5	23	13	9
New York	19	3	17	1	16
North Carolina	18	2	16	3	13
North Dakota	16	5	10	3	7
Ohio	15	4	11	2	9
Oklahoma	19	4	15	6	9
Oregon	—	—	—	—	—
Pennsylvania	18	3	15	4	11
Rhode Island	21	3	19	5	13
South Carolina	15	3	12	5	8
South Dakota	—	—	—	—	—
Tennessee	13	3	10	5	5
Texas	18	7	12	8	4
Utah	19	3	16	9	7
Vermont	22	4	18	5	13
Virginia	18	6	12	4	8
Washington	16	4	12	5	7
West Virginia	16	1	15	5	9
Wisconsin	18	4	14	3	11
Wyoming	16	3	13	5	8
Other jurisdictions					
District of Columbia	—	—	—	—	—
DoDEA ¹	11	2	9	3	6

— Not available. The jurisdiction did not participate or did not meet the minimum participation guidelines for reporting.

¹ Department of Defense Education Activity (overseas and domestic schools). Before 2005, DoDEA overseas and domestic schools were separate jurisdictions in NAEP. Pre-2005 data presented here were recalculated for comparability.

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), 1998, 2002, and 2007 Writing Assessments.

Table A-9

Eighth-grade public school students with disabilities identified, excluded, and assessed in NAEP writing, as a percentage of all students, by state: 1998, 2002, and 2007

State/jurisdiction	1998					2002				
	Identified	Excluded	Assessed	Assessed without accom- modations	Assessed with accom- modations	Identified	Excluded	Assessed	Assessed without accom- modations	Assessed with accom- modations
Nation (public)	11	4	8	5	3	13	3	10	5	5
Alabama	12	6	6	5	1	14	2	12	11	1
Alaska	—	—	—	—	—	—	—	—	—	—
Arizona	9	3	6	4	2	12	3	9	6	3
Arkansas	12	5	7	5	1	15	2	13	8	5
California	7	3	5	3	2	10	2	8	5	3
Colorado	9	3	6	4	2	—	—	—	—	—
Connecticut	14	6	8	5	3	15	3	11	5	6
Delaware	13	3	10	7	3	14	4	10	2	8
Florida	12	4	9	7	2	15	3	12	5	8
Georgia	10	4	6	4	2	11	3	8	4	4
Hawaii	10	3	7	5	2	15	2	13	7	6
Idaho	—	—	—	—	—	12	1	10	6	4
Illinois	10	3	6	4	2	13	2	11	5	7
Indiana	—	—	—	—	—	12	2	10	7	3
Iowa	—	—	—	—	—	—	—	—	—	—
Kansas	—	—	—	—	—	13	2	10	4	6
Kentucky	9	2	7	2	4	11	4	7	4	3
Louisiana	13	5	8	3	5	16	4	11	4	8
Maine	13	5	8	5	3	17	2	15	7	8
Maryland	12	2	10	4	7	14	3	11	8	3
Massachusetts	15	3	12	6	5	17	2	15	6	9
Michigan	—	—	—	—	—	13	5	8	3	4
Minnesota	11	2	8	6	2	13	2	11	7	4
Mississippi	9	5	5	4	1	10	5	5	3	2
Missouri	12	2	10	6	4	15	3	12	4	8
Montana	11	2	8	6	2	11	2	10	6	4
Nebraska	—	—	—	—	—	14	3	11	6	5
Nevada	11	4	7	4	2	13	3	11	6	5
New Hampshire	—	—	—	—	—	—	—	—	—	—
New Jersey	—	—	—	—	—	—	—	—	—	—
New Mexico	15	4	11	7	3	19	3	16	7	9
New York	9	2	8	2	6	15	4	11	3	8
North Carolina	12	3	9	3	6	16	4	11	3	8
North Dakota	—	—	—	—	—	13	1	13	7	5
Ohio	—	—	—	—	—	12	5	6	4	2
Oklahoma	12	8	4	3	1	14	2	12	8	4
Oregon	12	2	10	7	3	13	3	10	7	3
Pennsylvania	—	—	—	—	—	13	2	12	4	8
Rhode Island	14	3	10	8	2	17	2	15	6	9
South Carolina	12	5	7	5	2	14	5	9	5	4
South Dakota	—	—	—	—	—	—	—	—	—	—
Tennessee	12	4	8	7	1	14	3	11	9	2
Texas	14	5	9	7	2	13	5	8	7	1
Utah	8	3	5	4	1	11	2	9	6	4
Vermont	—	—	—	—	—	17	4	13	6	7
Virginia	12	4	9	5	3	15	5	10	4	6
Washington	10	2	7	5	2	11	2	9	4	5
West Virginia	14	5	9	5	3	18	4	14	5	9
Wisconsin	10	4	6	4	3	14	3	11	3	9
Wyoming	9	2	7	5	2	13	2	11	5	6
Other jurisdictions										
District of Columbia	10	5	5	4	1	17	5	12	4	8
DoDEA ¹	6	1	5	3	2	7	1	6	3	3

See notes at end of table.

Table A-9

Eighth-grade public school students with disabilities identified, excluded, and assessed in NAEP writing, as a percentage of all students, by state: 1998, 2002, and 2007—Continued

State/jurisdiction	2007				
	Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations
Nation (public)	13	3	10	3	8
Alabama	12	2	10	7	3
Alaska	—	—	—	—	—
Arizona	10	2	8	3	4
Arkansas	13	2	11	3	8
California	9	1	8	4	4
Colorado	9	2	7	1	6
Connecticut	11	1	10	2	8
Delaware	14	5	10	2	7
Florida	14	2	12	1	11
Georgia	12	2	9	2	7
Hawaii	13	1	13	4	8
Idaho	9	1	8	3	5
Illinois	14	2	12	2	10
Indiana	14	3	11	2	9
Iowa	15	2	13	2	11
Kansas	13	3	10	2	8
Kentucky	13	6	8	2	6
Louisiana	13	2	11	1	10
Maine	19	4	14	3	11
Maryland	—	—	—	—	—
Massachusetts	19	6	13	2	11
Michigan	14	4	11	2	9
Minnesota	11	2	10	3	7
Mississippi	10	2	8	1	7
Missouri	13	2	11	2	8
Montana	13	2	11	2	9
Nebraska	—	—	—	—	—
Nevada	12	2	10	5	6
New Hampshire	19	3	16	5	11
New Jersey	15	2	13	1	12
New Mexico	14	3	11	4	8
New York	16	2	13	1	13
North Carolina	15	2	13	2	11
North Dakota	15	5	9	3	7
Ohio	14	4	10	2	9
Oklahoma	16	4	12	4	9
Oregon	—	—	—	—	—
Pennsylvania	16	3	13	3	10
Rhode Island	18	1	17	4	13
South Carolina	13	3	10	4	7
South Dakota	—	—	—	—	—
Tennessee	12	3	8	3	5
Texas	12	6	7	3	3
Utah	9	2	7	2	6
Vermont	20	4	16	4	12
Virginia	14	5	9	3	7
Washington	12	3	8	2	6
West Virginia	15	1	14	5	9
Wisconsin	14	3	11	1	10
Wyoming	13	3	11	3	8
Other jurisdictions					
District of Columbia	—	—	—	—	—
DoDEA ¹	7	1	6	1	5

— Not available. The jurisdiction did not participate or did not meet the minimum participation guidelines for reporting.

¹ Department of Defense Education Activity (overseas and domestic schools). Before 2005, DoDEA overseas and domestic schools were separate jurisdictions in NAEP. Pre-2005 data presented here were recalculated for comparability.

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), 1998, 2002, and 2007 Writing Assessments.

Table A-10

Eighth-grade public school English language learners identified, excluded, and assessed in NAEP writing, as a percentage of all students, by state: 1998, 2002, and 2007

State/jurisdiction	1998					2002				
	Identified	Excluded	Assessed	Assessed without accom- modations	Assessed with accom- modations	Identified	Excluded	Assessed	Assessed without accom- modations	Assessed with accom- modations
Nation (public)	3	1	2	2	#	6	1	5	4	1
Alabama	#	#	#	#	#	1	#	1	#	#
Alaska	—	—	—	—	—	—	—	—	—	—
Arizona	10	3	7	6	1	13	3	10	9	1
Arkansas	1	1	#	#	#	3	1	2	1	#
California	17	4	13	13	#	21	2	19	17	1
Colorado	4	1	3	2	1	—	—	—	—	—
Connecticut	2	2	#	#	#	4	1	2	2	1
Delaware	1	#	1	1	#	2	1	1	#	#
Florida	4	1	3	2	#	7	2	6	3	3
Georgia	2	1	1	1	#	3	1	2	1	1
Hawaii	5	2	4	3	1	8	2	6	4	1
Idaho	—	—	—	—	—	4	1	3	2	#
Illinois	3	1	2	2	#	5	2	4	3	1
Indiana	—	—	—	—	—	1	1	1	#	#
Iowa	—	—	—	—	—	—	—	—	—	—
Kansas	—	—	—	—	—	4	1	3	2	1
Kentucky	1	#	#	#	#	#	#	#	#	#
Louisiana	1	#	#	#	#	1	#	1	1	#
Maine	1	#	#	#	#	2	#	2	1	#
Maryland	1	#	1	#	#	2	1	2	1	#
Massachusetts	2	2	1	#	#	4	2	3	2	1
Michigan	—	—	—	—	—	2	1	1	1	#
Minnesota	4	1	3	2	1	5	2	4	3	1
Mississippi	#	#	#	#	#	#	#	#	#	#
Missouri	1	#	#	#	#	1	#	1	1	#
Montana	#	#	#	#	#	3	#	2	2	#
Nebraska	—	—	—	—	—	3	1	2	1	1
Nevada	7	3	4	3	1	9	2	6	6	#
New Hampshire	—	—	—	—	—	—	—	—	—	—
New Jersey	—	—	—	—	—	—	—	—	—	—
New Mexico	10	3	7	7	1	20	3	17	13	3
New York	5	3	2	1	#	6	2	4	2	2
North Carolina	2	1	1	1	#	3	1	1	1	1
North Dakota	—	—	—	—	—	2	#	2	1	1
Ohio	—	—	—	—	—	1	#	1	1	#
Oklahoma	1	1	1	1	#	3	#	2	2	#
Oregon	2	1	2	1	#	7	1	5	5	1
Pennsylvania	—	—	—	—	—	1	#	1	#	#
Rhode Island	4	1	2	2	1	6	2	4	3	1
South Carolina	#	#	#	#	#	1	#	1	1	#
South Dakota	—	—	—	—	—	—	—	—	—	—
Tennessee	1	#	1	1	#	1	#	1	1	#
Texas	6	2	4	4	#	8	3	6	5	#
Utah	2	1	1	1	#	7	1	5	5	1
Vermont	—	—	—	—	—	1	#	1	1	#
Virginia	2	1	1	1	#	4	1	2	1	1
Washington	4	1	2	2	1	4	1	3	2	1
West Virginia	#	#	#	#	#	#	#	#	#	#
Wisconsin	2	1	1	1	#	4	2	2	1	1
Wyoming	#	#	#	#	#	2	#	2	2	#
Other jurisdictions										
District of Columbia	4	2	2	2	#	5	1	4	1	3
DoDEA ¹	2	1	1	1	#	5	2	3	3	1

See notes at end of table.

Table A-10

Eighth-grade public school English language learners identified, excluded, and assessed in NAEP writing, as a percentage of all students, by state: 1998, 2002, and 2007—Continued

State/jurisdiction	2007				
	Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations
Nation (public)	7	1	6	4	2
Alabama	2	#	1	1	#
Alaska	—	—	—	—	—
Arizona	10	1	9	7	2
Arkansas	4	#	4	1	2
California	21	1	20	17	2
Colorado	6	1	6	3	3
Connecticut	4	1	3	1	2
Delaware	2	1	1	#	1
Florida	6	1	5	1	4
Georgia	2	#	2	1	1
Hawaii	6	#	6	3	3
Idaho	6	1	5	4	1
Illinois	3	1	3	2	1
Indiana	3	1	2	1	1
Iowa	2	#	2	1	1
Kansas	4	1	4	2	1
Kentucky	1	#	1	1	#
Louisiana	1	#	1	#	1
Maine	2	1	2	1	1
Maryland	—	—	—	—	—
Massachusetts	4	1	3	1	2
Michigan	2	#	2	1	1
Minnesota	6	1	5	4	2
Mississippi	1	#	1	#	#
Missouri	2	#	2	1	1
Montana	4	#	4	2	2
Nebraska	—	—	—	—	—
Nevada	11	2	9	7	2
New Hampshire	2	#	1	1	1
New Jersey	3	1	2	1	1
New Mexico	17	3	14	11	3
New York	5	1	4	#	4
North Carolina	4	#	4	1	2
North Dakota	2	#	2	1	1
Ohio	1	#	1	#	1
Oklahoma	3	#	3	2	1
Oregon	—	—	—	—	—
Pennsylvania	2	1	2	1	1
Rhode Island	4	1	3	2	1
South Carolina	2	#	2	1	1
South Dakota	—	—	—	—	—
Tennessee	2	#	2	1	1
Texas	8	2	6	4	2
Utah	10	1	9	7	2
Vermont	2	#	2	1	1
Virginia	4	1	3	2	1
Washington	6	1	4	3	2
West Virginia	1	#	1	1	#
Wisconsin	5	1	3	1	2
Wyoming	3	#	3	2	1
Other jurisdictions					
District of Columbia	—	—	—	—	—
DoDEA ¹	4	1	3	2	1

— Not available. The jurisdiction did not participate or did not meet the minimum participation guidelines for reporting.

Rounds to zero.

¹ Department of Defense Education Activity (overseas and domestic schools). Before 2005, DoDEA overseas and domestic schools were separate jurisdictions in NAEP. Pre-2005 data presented here were recalculated for comparability.

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), 1998, 2002, and 2007 Writing Assessments.

Table A-11

Eighth-grade public school students with disabilities (SD) and/or English language learners (ELL) identified, excluded, and assessed in NAEP writing, as a percentage of all students, by SD/ELL category and urban district: 2002 and 2007

SD/ELL category and district	2002					2007				
	Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations	Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations
SD and/or ELL										
Nation (public)	18	4	14	8	5	18	3	15	6	9
Large central city (public)	23	5	19	14	5	24	4	20	10	10
Atlanta	8	3	5	4	1	11	2	10	2	7
Austin	—	—	—	—	—	27	6	22	16	6
Boston	—	—	—	—	—	28	6	22	6	16
Charlotte	—	—	—	—	—	19	3	16	6	10
Chicago	24	7	17	10	7	23	5	17	4	13
Cleveland	—	—	—	—	—	24	11	13	2	11
District of Columbia	21	6	15	5	10	—	—	—	—	—
Houston	27	8	20	20	#	22	8	14	11	4
Los Angeles	35	5	31	27	4	34	2	31	24	7
New York City	27	7	20	9	11	23	2	21	2	19
San Diego	—	—	—	—	—	28	3	24	18	6
SD										
Nation (public)	13	3	10	5	5	13	3	10	3	8
Large central city (public)	13	3	11	6	4	13	3	10	3	7
Atlanta	7	3	4	4	1	10	2	8	2	6
Austin	—	—	—	—	—	16	4	12	7	5
Boston	—	—	—	—	—	19	5	14	2	12
Charlotte	—	—	—	—	—	12	2	10	2	8
Chicago	18	3	14	8	7	18	3	14	2	12
Cleveland	—	—	—	—	—	20	10	10	1	9
District of Columbia	17	5	12	4	8	—	—	—	—	—
Houston	15	5	10	10	#	12	5	7	3	3
Los Angeles	13	2	11	8	3	10	2	9	3	5
New York City	17	3	14	6	9	14	1	13	1	12
San Diego	—	—	—	—	—	11	3	8	3	5
ELL										
Nation (public)	6	1	5	4	1	7	1	6	4	2
Large central city (public)	13	3	10	9	1	12	2	11	7	3
Atlanta	1	1	1	1	#	2	#	2	1	1
Austin	—	—	—	—	—	14	3	11	10	1
Boston	—	—	—	—	—	12	3	9	4	4
Charlotte	—	—	—	—	—	8	1	7	4	3
Chicago	8	4	4	3	1	7	3	4	2	2
Cleveland	—	—	—	—	—	5	2	3	1	2
District of Columbia	5	1	4	1	3	—	—	—	—	—
Houston	18	5	14	14	#	13	4	9	8	1
Los Angeles	30	4	26	24	2	28	2	27	22	4
New York City	14	6	8	5	3	12	2	10	1	9
San Diego	—	—	—	—	—	20	1	19	16	3

— Not available. The district did not participate.

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. As of 2005, "large central city" includes nationally representative public schools located in large central cities (population of 250,000 or more) within a Metropolitan Statistical Area (MSA). 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), 2002 and 2007 Trial Urban District Writing Assessments.

Investigating the Potential Effects of Exclusion Rates on Assessment Results

Variation in the rates of exclusion of students with disabilities (SD) and English language learners (ELL) introduces validity concerns for comparisons over time or between jurisdictions. The essential problem is the differential representativeness of samples, which could impact the comparability of cross-state comparisons within a given year and state trends across years. Because students with disabilities and English language learners tend to score below average on assessments, excluding such students may increase a jurisdiction's scores. Conversely, including more of these students might depress score gains. In 2007, exclusion rates varied among jurisdictions. In addition, cases of both increases and decreases in exclusion rates occurred between 1998 and 2007, making comparisons over time within jurisdictions complex to interpret. Tables A-8 and A-11 on the preceding pages display the rates of exclusion in each jurisdiction for grade 8.

One factor that contributed to the variability in exclusion rates across states is that the percentage of students who are *identified* as having disabilities or as English language learners varies across jurisdictions. Some reasons for the variation include lack of standardized criteria for defining students as having specific disabilities or as ELL, the wide differences in percentages of immigrant children across states, and changes or differences in policy and practices regarding implementation of the Individuals with Disabilities Education Act (IDEA).

Types of Accommodations Permitted

Table A-12 displays the percentages of SD and ELL students assessed with the available accommodations. It should be noted that students assessed with accommodations typically received some combination of accommodations. The numbers and percentages presented in the table reflect only the primary accommodation provided. For example, students assessed in small groups (as compared with standard NAEP sessions of about 30 students) usually received extended time. Here, the primary accommodation coded would be small groups. In one-on-one administrations, students often received assistance in recording answers (e.g., use of a scribe or computer) and were afforded extra time. Extended time was considered the primary accommodation only when it was the sole accommodation provided. The assessment did not allow some accommodations that were permitted in certain states in past assessments. Some states have allowed questions and, in some cases, reading passages to be read aloud to the students. In designing the reading assessment, reading aloud as an accommodation was viewed as changing the nature of the construct being measured and, hence, was not permitted. Because NAEP considers the domain of its reading assessment to be reading in English, no attempt was made to provide an alternate language version of the assessment, and the use of bilingual dictionaries was not permitted. In the writing assessment, however, reading the essay prompts aloud was permitted because it did not change the construct being measured.

Table A-12

Percentage of eighth- and twelfth-grade public and nonpublic school students identified as students with disabilities (SD) and/or English language learners (ELL) assessed in NAEP writing with accommodations, by SD/ELL category and type of primary accommodation: 1998, 2002, and 2007

SD/ELL category and type of accommodation	Grade 8			Grade 12		
	1998	2002	2007	1998	2002	2007
SD and/or ELL						
Bilingual dictionary	#	0.1	0.4	#	0.1	0.3
Large-print book	#	#	#	#	#	#
Extended time	0.9	1.9	3.8	0.5	1.4	2.9
Read aloud	0.1	0.3	2.6	#	0.2	1.8
Small group	1.7	2.7	0.9	0.7	1.1	0.5
One-on-one	0.1	0.1	#	0.1	0.1	#
Scribe/computer	0.1	#	0.2	#	#	0.3
Breaks	—	—	#	—	—	#
Magnifying device	—	—	#	—	—	#
School staff administers	—	—	#	—	—	#
Other	#	0.1	0.5	#	#	0.2
SD						
Bilingual dictionary	#	#	#	#	#	#
Large-print book	#	#	#	#	#	#
Extended time	0.7	1.7	3.4	0.4	1.3	2.7
Read aloud	0.1	0.3	2.4	#	0.2	1.7
Small group	1.6	2.6	0.8	0.7	1.1	0.4
One-on-one	0.1	0.1	#	0.1	0.1	#
Scribe/computer	0.1	#	0.2	#	#	0.3
Breaks	—	—	#	—	—	#
Magnifying device	—	—	#	—	—	#
School staff administers	—	—	#	—	—	#
Other	#	0.1	0.5	#	#	0.2
ELL						
Bilingual dictionary	#	0.1	0.4	#	0.1	0.3
Large-print book	#	#	#	#	#	#
Extended time	0.1	0.4	0.7	0.1	0.1	0.4
Read aloud	#	#	0.3	#	#	0.2
Small group	0.1	0.3	0.1	#	#	0.1
One-on-one	#	#	#	#	#	#
Scribe/computer	#	#	#	#	#	#
Breaks	—	—	#	—	—	#
Magnifying device	—	—	#	—	—	#
School staff administers	—	—	#	—	—	#
Other	#	#	0.1	#	#	#

— Not available.

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), 1998, 2002, and 2007 Writing Assessments.

Data Collection and Scoring

The 2007 NAEP writing assessment was conducted from January to March 2007 by contractors to the U.S. Department of Education. Trained field staff from Westat conducted the data collection. Materials from the 2007 assessment were shipped to Pearson, Inc., where the test booklets and surveys were scanned.

Trained staff evaluated the responses to the essay questions using scoring rubrics or guides prepared by Educational Testing Service (ETS). Each essay prompt had a unique scoring guide that defined the criteria used to evaluate students' responses. All the writing tasks were evaluated according to a six-level scoring guide. At each grade, scoring guides were developed for each of the three types of writing tasks: narrative, informative, and persuasive.

Specialists in writing who are highly experienced in teaching and/or assessing writing trained the professional raters who evaluated the student responses. The trainers received extensive training together that included reading a manual that explained how to use the scoring guides and the processes for training and checking raters. For each task, the trainer, in consultation with other trainers or assessment specialists, chose numerous sample responses to present to raters and prepared notes on how the scoring guide applied to the particular task. The sample responses helped raters become accustomed to the variety of responses the task elicited before they began rating the student responses. Raters had to pass a qualifying test before they could evaluate student responses: they had to agree with at least 70 percent of the ratings (to a set of 10 student responses) that were given beforehand by their trainer.

In order to determine interrater reliability of scoring, a specified percentage of responses was read twice: two raters read 6 percent of the responses at grade 8 (where there was the large state-level sample) and 25 percent of responses at grade 12 (national sample only).

For the national and state writing assessments, approximately 356,132 responses to writing tasks were scored in 2007. This number includes rescoring to monitor interrater reliability. Like other NAEP subjects, the percentage of exact agreement of ratings was used as an interrater reliability measure in the writing assessment. The within-year percentage of exact agreement of ratings on the six-level scoring guides for the 2007 reliability samples was 72 percent at grade 8 and 65 percent at grade 12. The acceptable level of the percentage of exact agreement of ratings for a 6-category item in NAEP writing is 60 percent.

Data Analysis and IRT Scaling

After the professional scoring, all information was transcribed into the NAEP database at ETS. Each processing activity was conducted with rigorous quality control. After the assessment information was compiled in the database, the data were weighted according to the population structure. The weighting for the national and state samples reflected the probability of selection for each student as a result of the sampling design, adjusted for nonresponse.⁴

Analyses were then conducted to determine the percentages of students who wrote responses to each writing task at each level on the scoring guide and who provided various responses to each background question. In calculating response percentages for each task, only students classified as having been presented the question were included in the denominator of the statistic. Students whose papers were blank or whose responses were judged to be off topic were similarly excluded from the calculation of the scale.

Item Response Theory (IRT) was used to estimate average writing scale scores for the nation, for various student groups of interest within the nation, and for the states, other jurisdictions, and trial urban districts. 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 in NAEP is to provide a common scale on which performance can be compared among groups, such as those defined by characteristics including gender and race/ethnicity, even when students receive different writing tasks. One desirable feature of IRT is that it locates items and students on this common scale. In contrast to classical test theory, IRT does not rely solely on the total number of correct item responses, but uses the particular patterns of student responses to items in determining the student location on the scale. As a result, adding items that function at a particular point on the scale to the assessment does not change the location of the students on the scale, even though students may respond correctly to more items. It does increase the relative precision with which students are measured, particularly those students whose scale locations are close to the additional items.

The results for 1998, 2002, and 2007 are presented on the NAEP writing scale (ranging from 0 to 300). Developed for the 1998 assessment, the scale was computed to report performance at each grade level. The scale summarizes student performance across all three purposes for writing (narrative, informative and persuasive) in the assessment. There were not enough writing tasks to create a separate subscale for each writing purpose.

In producing the writing scale, an IRT model was used. The writing tasks (all rated according to six-level scoring guides) were scaled by use of a Generalized Partial-Credit (GPC) model.⁵ First used in 1992, the GPC model permits the scaling of questions scored according to multipoint rating schemes. The model takes full advantage of the information available from each of the student response categories used for more complex constructed-response questions such as writing tasks.⁶

Because the NAEP design gives each student a small proportion of the pool of assessment items, the assessment cannot provide reliable information about individual performance. Traditional test scores for individual students, even those based on IRT, would result in misleading estimates of population characteristics, such as 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 directly, without the intermediary computation of estimates of individuals. This is accomplished using marginal estimation scaling model techniques for latent variables.⁷ Under the assumptions of the scaling models, these population estimates will be consistent in the sense that the estimates approach the model-based population values as the sample size increases. This would not be the case for population estimates obtained by aggregating optimal estimates of individual performance.⁸

Weighting and Variance Estimation

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

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

Not only must appropriate estimates of population characteristics be derived, but appropriate measures of the degree of uncertainty must be obtained for those statistics. Two components of uncertainty are accounted for in the variability of statistics based on student ability: the uncertainty due to sampling only a relatively small number of students and the uncertainty due to sampling only a portion of the cognitive domain of interest (in this case, writing). The first component accounts for the variability associated with the estimated percentages of students who had certain background characteristics or who had a certain rating for their responses to a task.

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 two writing tasks, the scale score for any single student would be imprecise. In this case, NAEP's marginal estimation methodology can be used to describe the performance of different groups of students. The estimate of the variance of the students' posterior 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.⁹

In some circumstances, it is not possible to obtain appropriate estimates of standard errors, and the accuracy of the statistic being estimated may then be called into question. In the case of extreme percentages, close to 100 or 0 for student group percentages and percentages at or above achievement levels, the standard error may have unknown accuracy or be undefined. In such cases, tables of NAEP results in the NAEP Data Explorer software tool display the symbol *** in place of the standard error and provide the notation: Standard error cannot be determined.

When a standard error is based on a small number of students, or the group of students is enrolled in a small number of schools, the amount of uncertainty associated with the estimation of the standard error may be quite large, and the accuracy of both the standard error and the estimate of the statistic are compromised. An indicator that is used in NAEP for these situations is the "rule of five." The rule of five requires that estimates of statistics be based on at least five sampling units (e.g., schools). If the requirement is not met, tables of NAEP results insert the symbol ‡ in place of both the statistic and its standard error, and provide the notation: Reporting standards not met.

The symbol ‡ and its accompanying notation are also used in other instances. For example, it is used when the sample size falls below the minimum of 62 students needed to ensure enough power to detect certain effects, and when response rates fall below certain levels. However, these instances are largely unrelated to concerns about weighting or variance estimation.

The reader is reminded that, as with findings from all surveys, NAEP results are subject to other kinds of error, including the effects of imperfect adjustment 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 background information; mistakes in recording, coding, or scoring data; and other errors in collecting, processing, sampling, and estimating missing data. The extent of nonsampling errors is assumed to be small but 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.

Drawing Inferences From the Results

The reported statistics are estimates and are therefore subject to a measure of uncertainty. There are two sources of such uncertainty. First, NAEP uses a sample of students rather than testing all students. Second, all assessments have some amount of uncertainty because they cannot ask all the questions that might be asked in a content area. The magnitude of this uncertainty is reflected in the standard error of each of the estimates. When the percentages or average scale scores of certain groups are compared, the estimated standard error should be taken into account. Therefore, the comparisons are based on statistical tests that consider the estimated standard errors of those statistics and the magnitude of the difference among the averages or percentages.

For the data in this report, all the estimates have corresponding estimated standard errors of the estimates. For example, tables A-13 and A-14 show the average national scale score for the NAEP 1998, 2002, and 2007 national assessments and the percentage of students within each achievement-level range and at or above achievement levels. In both tables, estimated standard errors appear in parentheses next to each estimated scale score or percentage. For the estimated standard errors corresponding to other data in this report, the reader can go to the NAEP Data Explorer tool on the NCES website at <http://nces.ed.gov/nationsreportcard/naepdata>.

Using confidence intervals based on the standard errors provides a way to take into account the uncertainty associated with sample estimates and to make inferences about the population averages and percentages in a manner that reflects that uncertainty. An estimated sample average scale score plus or minus 1.96 standard errors approximates a 95 percent confidence interval for the corresponding population quantity. This statement means that one can conclude with an approximately 95 percent level of confidence that the average performance of the entire population of interest (e.g., all fourth-grade students in public and nonpublic schools) is within plus or minus 1.96 standard errors of the sample average.

For example, suppose that the average writing scale score of the students in a particular group was 256 with an estimated standard error of 1.2. An approximately 95 percent confidence interval for the population quantity would be as follows:

$$\begin{aligned} &\text{Average} \pm 1.96 \text{ standard errors} \\ &= 256 \pm 1.96 \times 1.2 \\ &= 256 \pm 2.4 \end{aligned}$$

Therefore, the 95% confidence interval is bounded by: (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. It should be noted that this example and the examples in the following sections are illustrative. More precise estimates carried out to one or more decimal places are used in the actual analyses.

Similar symmetric confidence intervals can be constructed for percentages, if the percentages are not extremely large or small. For extreme percentages, a symmetric interval based on a normal distribution is not appropriate, and the common standard error calculation is possibly problematic. Standard errors of extreme percentages should be interpreted with caution.

Table A-13**Average scale scores and standard errors for public and nonpublic school students in NAEP writing, by grade: 1998, 2002, and 2007**

Grade	1998	2002	2007
Grade 8	150 (0.6) *	153 (0.5) *	156 (0.2)
Grade 12	150 (0.7) *	148 (0.8) *	153 (0.6)

* Significantly different from the score in 2007.

NOTE: Standard errors of the estimated scale scores appear in parentheses. Beginning in 2002, NAEP sample sizes at grade 8 have increased compared to previous years, resulting in smaller detectable differences than in previous assessments.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998, 2002, and 2007 Writing Assessments.

Table A-14**Percentage of public and nonpublic school students and standard errors in NAEP writing, by achievement-level performance, grade, and assessment year: 1998, 2002, and 2007**

Grade and year	Below <i>Basic</i>	At <i>Basic</i>	At <i>Proficient</i>	At <i>Advanced</i>	At or above <i>Basic</i>	At or above <i>Proficient</i>
Grade 8						
1998	16 (0.5) *	58 (0.5) *	25 (0.7) *	1 (0.1) *	84 (0.5) *	27 (0.7) *
2002	15 (0.4) *	54 (0.5) *	29 (0.5) *	2 (0.1)	85 (0.4) *	31 (0.6)
2007	12 (0.2)	56 (0.3)	31 (0.2)	2 (0.1)	88 (0.2)	33 (0.3)
Grade 12						
1998	22 (0.7) *	57 (0.7)	21 (0.7) *	1 (0.1)	78 (0.7) *	22 (0.7) *
2002	26 (0.7) *	51 (0.7) *	22 (0.7)	2 (0.2) *	74 (0.7) *	24 (0.8)
2007	18 (0.5)	57 (0.4)	23 (0.5)	1 (0.1)	82 (0.5)	24 (0.6)

* Significantly different from the percentage in 2007.

NOTE: Standard errors of the estimated percentages appear in parentheses. Beginning in 2002, NAEP sample sizes at grade 8 have increased compared to previous years, resulting in smaller detectable differences than in previous assessments. 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), 1998, 2002, and 2007 Writing Assessments.

Analyzing Group Differences in Averages and Percentages

Statistical tests determine whether, based on the data from the groups in the sample, there is strong enough evidence 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 or lower than another group), regardless of whether the sample averages or percentages appear to be approximately the same. 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.

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

The standard error of the difference can be used, just like the standard error for an individual group average or percentage, to help determine whether differences among groups in the population are real. The difference between the averages or percentages of the two groups plus or minus 1.96 standard errors of the difference represents an approximately 95 percent confidence interval. If the resulting interval includes zero, there is insufficient evidence to claim a real difference between the groups in the population. If the interval does not contain zero, the difference between the groups is statistically significant at the .05 level.

The following example of comparing groups addresses the problem of determining whether the average writing scale score of group A is higher than that of group B. The sample estimates of the average scale scores and estimated standard errors are 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 two points (218 - 216). The estimated standard error of this difference is

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

Thus, an approximately 95 percent confidence interval for this difference is plus or minus 1.96 standard errors of the difference.

$$2 \pm 1.96 \times 1.4$$

$$2 \pm 2.7$$

$$(-0.7, 4.7)$$

The value zero is within the confidence interval; therefore, there is insufficient evidence to conclude that group A performed statistically differently from group B.

The procedure above is appropriate to use when it is reasonable to assume that the groups being compared have been independently sampled for the assessment. Such an assumption is clearly warranted when comparing results across assessment years (e.g., comparing the 2002 and 2007 results for a particular state or group) or when comparing results for one state with another. This is the approach used for NAEP reports when comparisons involving independent groups are made. The assumption of independence is violated to some degree when comparing group results for the nation or a particular state (e.g., comparing national 2007 results for males and females), since these samples of students have been drawn from the same schools. When the groups being compared do not share students (as is the case, for example, when comparing males and females), the impact of this violation of the independence assumption on the outcome of the statistical tests is assumed to be small, and NAEP, by convention, has, for computational convenience, routinely applied the procedures described above to those cases as well.

When making comparisons of results for groups that share a considerable proportion of students in common, it is not appropriate to ignore such dependencies. In such cases, NAEP has used procedures appropriate to comparing dependent groups. When the dependence in group results is due to the overlap in samples (e.g., when a subgroup is being compared to a total group), a simple modification of the usual standard error of the difference formula can be used. The formula for such cases is:

$$SE_{\text{Total-Subgroup}} = \sqrt{(SE_{\text{Total}}^2 + SE_{\text{Subgroup}}^2 - 2pSE_{\text{Subgroup}}^2)}$$

where p is the proportion of the total group contained in the subgroup.¹⁰ This formula was used for this report when a state was compared to the aggregate nation.

Conducting Multiple Tests

The procedures used to determine whether group differences in the samples represent actual differences among the groups in the population 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, there are times when 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., .05), standard methods must be adjusted by multiple comparison procedures.¹¹ One such procedure, the Benjamini-Hochberg False Discovery Rate (FDR) procedure, was used to control the certainty level.¹²

Unlike other multiple comparison procedures 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. (A "family" in this context is the number of categories to be compared for a given variable. This might be six within the race/ethnicity variable or 50 when considering states.) Furthermore, the FDR procedure used in NAEP is considered appropriately less conservative than familywise procedures for large families of comparisons.¹³ Therefore, the FDR procedure is more suitable for multiple comparisons in NAEP than other procedures.

To illustrate how the FDR procedure is used, consider the comparisons of current and previous years' average scale scores for the five groups presented in table A-15. Note that the difference in average scale scores and the estimated standard error of the difference are calculated as the example in the previous section. The test statistic shown is the difference in average scale scores divided by the estimated standard error of the difference. (Rounding of the data occurs after the test is done.)

Table A-15

Example of False Discovery Rate comparisons of average scale scores for different groups of students

Group	Previous year		Current year		Previous year and current year			
	Average scale score	Standard error	Average scale score	Standard error	Differences in averages	Standard error of differences	Test statistic	Percent confidence ¹
1	224	1.3	226	1.0	2.08	1.62	1.29	20
2	187	1.7	193	1.7	6.31	2.36	2.68	1
3	191	2.6	197	1.7	6.63	3.08	2.15	4
4	229	4.4	232	4.6	3.24	6.35	0.51	62
5	201	3.4	196	4.7	-5.51	5.81	-0.95	35

¹ The percent confidence is $2(1-F(x))$, where $F(x)$ is the cumulative distribution of the t -distribution with the degrees of freedom adjusted to reflect the complexities of the sample design.

NOTE: Data in table are for illustration purposes only and are not actual NAEP data.

The difference in average scale scores and its estimated standard error can be used to find an approximately 95 percent confidence interval, or they can be used to identify a confidence percentage. The confidence percentage for the test statistics is identified from statistical tables instead of checking to see whether zero is within the 95 percent confidence interval about the mean. The significance level from the statistical tables can be directly compared to the maximum acceptable error of 5 percent ($100 - 95 = 5$ percent).

If the comparison of average scale scores across two years were made for only one of the five groups, there would be a significant difference between the average scale scores for the two years at a significance level of less than 5 percent. However, because of interest in the difference in average scale scores across the two years for all five of the groups, comparing each of the significance levels to 5 percent is not adequate. Groups of students defined by shared characteristics, such as racial/ethnic groups, are treated as sets or families when making comparisons. However, comparisons of average scale scores for each pair of years were treated separately, so the steps described in this example would be replicated for the comparison of other current and previous year average scale scores.

Using the FDR procedure to take into account that all comparisons are of interest, the percents of confidence in the example are ordered from largest to smallest: 62, 35, 20, 4, and 1. In the FDR procedure, 62 percent confidence for the group 4 comparison would be compared to 5 percent, 35 percent for the group 5 comparison would be compared to $0.05 \times (5-1)/5 = 0.04 = 4$ percent,¹⁴ 20 percent for the group 1 comparison would be compared to $0.05 \times (5-2)/5 = 0.03 = 3$ percent, 4 percent for the group 3 comparison would be compared to $0.05 \times (5-3)/5 = 0.02 = 2$ percent, and 1 percent for the group 2 comparison (actually slightly smaller than 1 prior to rounding) would be compared to $0.05 \times (5-4)/5 = 0.01 = 1$ percent. The procedure stops with the first contrast found to be significant. The last of these comparisons is the only one for which the percent confidence is smaller than the FDR procedure value. The difference between the current year's and previous years' average scale scores for the group 2 students is significant; for all of the other groups, average scale scores for the current and previous year are not significantly different from one another. In practice, a very small number of counterintuitive results occur when the FDR procedures are used to examine between-year differences in subgroup results by jurisdiction. In those cases, results were not included in this report.

Understanding NAEP Reporting Groups

NAEP results are provided for groups of students defined by shared characteristics—gender, race/ethnicity, parental education, region of the country, type of school, school's type of location (categorized by population density), and eligibility for free/reduced-price school lunch under the National School Lunch Program. Based on participation rate criteria, results are reported for subpopulations only when sufficient numbers of students and adequate school representation are present. In addition, based on statistical considerations about power and variance estimation, the minimum requirement on which to base any statistic is at least 62 students in a particular subgroup from at least five primary sampling units (PSUs).¹⁵ Definitions of the subpopulations are presented below.

Gender: Results are reported separately for male and female students.

Race/Ethnicity: In all NAEP assessments, data about student race/ethnicity are collected from two sources: school records and student self-reports. Prior to 2002, NAEP used students' self-reported race as the primary race/ethnicity reporting variable. Beginning in 2002, the race/ethnicity variable presented in NAEP reports has been based on the race reported by the school. When school-recorded information is missing, student-reported data are used to determine race/ethnicity. Therefore, beginning in 2002 the data for racial/ethnic groups included for all assessment years are based on the school-reported race/ethnicity variable. Information on student race/ethnicity is reported as one of six categories: White, Black, Hispanic, Asian/Pacific Islander, American Indian/Alaska Native, and Unclassified. Black includes African American, Hispanic includes Latino, and Pacific Islander includes Native Hawaiian. Race categories exclude Hispanic origin unless specified. Unclassified students are those whose school-reported race/ethnicity was "other" or "unavailable" or was missing, and whose race/ethnicity category could not be determined from self-reported information. Information based on student self-reported race/ethnicity is available on the NAEP Data Tool (<http://nces.ed.gov/nationsreportcard/naepdata/>).

Parental Education: Eighth- and twelfth-graders 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. Prior to 2005, parental education questions were presented to students at grade 4, but were not reported because their responses were highly variable. Starting in 2005, parental education questions were not presented to students at grade 4.

Region of the Country: Prior to 2003, NAEP results were reported for four NAEP-defined regions of the nation: Northeast, Southeast, Central, and West. As of 2003, to align NAEP with other federal data collections, NAEP analysis and reports have used the U.S. Census Bureau's definition of "region." The four regions defined by the U.S. Census Bureau are Northeast, South, Midwest, and West. The Central region used by NAEP before 2003 contained the same states as the Midwest region defined by the U.S. Census. The former Southeast region consisted of the states in the Census-defined South minus Delaware, the District of Columbia, Maryland, Oklahoma, Texas, and the section of Virginia in the District of Columbia metropolitan area. The former West region consisted of Oklahoma, Texas, and the states in the Census-defined West. The former Northeast region consisted of the states in the Census-defined Northeast plus Delaware, the District of Columbia, Maryland, and the section of Virginia in the District of Columbia metropolitan area. Therefore, trend data by region are provided in NAEP reports for 2003 and 2005 only. Figure A-2 shows how states are subdivided into these census regions. All 50 states and the District of Columbia are listed. The Department of Defense Education Activity schools are not assigned to any region.

Figure A-2

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.

Type of School: Results are reported by the type of school that the student attends—public or private. Private schools include Catholic and other private schools.¹⁶ Because they are funded by federal authorities (not state/local governments), Bureau of Indian Education (BIE) schools and Department of Defense Education Activity schools (DoDEA) are not included in either the public or private categories; they are included in the overall national results. State-level reporting in NAEP includes only public schools. The national sample reporting for NAEP includes public, private, the DoDEA, and BIE schools.

Type of Location: NAEP results are reported for four mutually exclusive categories of school location: 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. The classification system was revised for 2007; therefore, trend comparisons to previous years are not available. The new locale codes are based on an address's proximity to an urbanized area (a densely settled core with densely settled surrounding areas). This is a change from the original system based on metropolitan statistical areas. To distinguish the two systems, the new system is referred to as "urban-centric locale codes."

The urban-centric locale code system classifies territory into four major types: city, suburban, town, and rural. Each type has three subcategories. For city and suburb, these are gradations of size—large, midsize, and small. Towns and rural areas are further distinguished by their distance from an urbanized area. They can be characterized as fringe, distant, or remote. More detail on the locale codes is available at http://nces.ed.gov/ccd/rural_locale.asp.

Eligibility for the National School Lunch Program: As part of the Department of Agriculture's National School Lunch Program, schools can receive cash subsidies and donated commodities in turn for offering free or reduced-price lunches to eligible children. Based on available school records, students were classified as either currently eligible for free/reduced-price school lunch or not eligible. 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, 2006 through June 30, 2007, for a family of four, 130 percent of the poverty level was \$26,000, and 185 percent was \$37,000.) Additional information on eligibility may be found at the Department of Agriculture website at <http://www.fns.usda.gov/cnd/lunch/>. The classification applies only to the school year when the assessment was administered (i.e., the 2006–07 school year) and is not based on eligibility in previous years. If school records were not available, the student was 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." The percentage of students in this category has declined in recent assessments. The decline has sometimes been sufficiently large as to preclude the reporting of trend data. Real changes in the eligible student percentages could not be distinguished from changes resulting from improved data collection methods.

End Notes

¹ Section 504 of the Rehabilitation Act of 1973 is a civil rights law designed to prohibit discrimination on the basis of disability in programs and activities, including education, that receive federal financial assistance.

² Office of Special Education Programs. (1997). *To Assure the Free Appropriate Public Education of all Children with Disabilities. Nineteenth Annual Report to Congress on the Implementation of the Individuals With Disabilities Education Act*. Archived at the U.S. Department of Education website: <http://www.ed.gov/about/offices/list/osers/index.html>.

³ The two samples are described as "overlapping" because, in 1998 and 2000, the same group of non-SD and/or ELL students was included in both samples.

⁴ Weighting procedures are described more fully in the "Weighting and Variance Estimation" section in this document.

⁵ Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological*

Measurement, 16(2): 159–176.

⁶ More detailed information regarding the IRT analyses used in NAEP will be included in the technical documentation section of the NAEP website (<http://nces.ed.gov/nationsreportcard>).

⁷ Mislevy, R.J. and Sheehan, K.M. (1987). Marginal Estimation Procedures. In A.E. Beaton (Ed.), *Implementing the New Design: The NAEP 1983–1984 Technical Report* (Technical Rep. No. 15-TR-20), pp. 293–360. Princeton, NJ: Educational Testing Service.

⁸ For theoretical and empirical justification of the procedures employed, see Mislevy, R.J. (1988). Randomization-Based Inferences About Latent Variables From Complex Samples. *Psychometrika*, 56(2): 177–196.

⁹ For further details, see Johnson, E.G., and Rust, K.F. (1992). Population Inferences and Variance Estimation for NAEP Data. *Journal of Educational Statistics*, 17(2): 175–190.

¹⁰ This is a special form of the common formula for standard error of dependent samples. The standard formula can be found, for example, in Kish, L. (1995). *Survey Sampling*. New York: John Wiley and Sons, Inc.

¹¹ Miller, R.G. (1981). *Simultaneous Statistical Inference* (2nd ed.). New York: Springer-Verlag.

¹² Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B*, (1): 289–300.

¹³ Williams, V.S.L., Jones, L.V., and Tukey, J.W. (1999). Controlling Error in Multiple Comparisons with Examples From State-to-State Differences in Educational Achievement. *Journal of Educational and Behavioral Statistics*, 24(1): 42–69.

¹⁴ The level of confidence times the number of comparisons minus one divided by the number of comparisons is $0.05 \times (5-1)/5 = 0.04 = 4$ percent.

¹⁵ For the NAEP national assessments prior to 2002, a PSU is a selected geographic region (a county, group of counties, or metropolitan statistical area). Since 2002, the first-stage sampling units are schools (public and nonpublic) in the selection of the combined sample. Further details about the procedure for determining minimum sample size will appear in the technical documentation section of the NAEP website at <http://nces.ed.gov/nationsreportcard>.

¹⁶ A more detailed breakdown of private school results is available on the NAEP website at <http://nces.ed.gov/nationsreportcard/naepdata>.

What is the Nation's Report Card™?

The Nation's Report Card informs the public about the academic achievement of elementary and secondary students in the United States. Report cards communicate the findings of the National Assessment of Educational Progress (NAEP), the only continuing and nationally representative measure of achievement in various subjects over time. *The Nation's Report Card* compares performance among states, urban districts, public and private schools, and student demographic groups.

For over three decades, NAEP assessments have been conducted periodically in reading, mathematics, science, writing, history, geography, and other subjects. By making objective information available on student performance 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 and relevant variables is collected. The privacy of individual students is protected, and the identities of participating schools are not released.

NAEP is a congressionally authorized project of the National Center for Education Statistics within the Institute of Education Sciences of the U.S. Department of Education. By law, the Commissioner of Education Statistics is responsible for carrying out the NAEP project. The National Assessment Governing Board oversees and sets policy for NAEP. The Governing Board is an independent, bipartisan group whose members include governors, state legislators, local and state officials, educators, business representatives, and members of the general public. The Governing Board's mission is, "to ensure equal access to education and to promote educational excellence throughout the nation."

The National Assessment Governing Board

Darvin M. Winick, Chair

President
Winick & Associates
Austin, Texas

Amanda P. Avallone, Vice Chair

Assistant Principal and Eighth-Grade Teacher
Summit Middle School
Boulder, Colorado

Francie Alexander

Chief Academic Officer
Scholastic, Inc.
Senior Vice President
Scholastic Education
New York, New York

David J. Alukonis

Chairman
Hudson School Board
Hudson, New Hampshire

Gregory Cizek

Professor of Educational Measurement
University of North Carolina
Chapel Hill, North Carolina

Honorable David P. Driscoll

Former Commissioner of Education
Massachusetts Department of Education
Malden, Massachusetts

Louis M. Fabrizio

Director, Division of Accountability Services
North Carolina Department of Public Instruction
Raleigh, North Carolina

Alan J. Friedman

Consultant
Museum Development and Science Communication
New York, New York

David W. Gordon

County Superintendent of Schools
Sacramento County Office of Education
Sacramento, California

Robin C. Hall

Principal
Beecher Hills Elementary School
Atlanta, Georgia

Kathi M. King

Twelfth-Grade Teacher
Messalonskee High School
Oakland, Maine

Honorable Keith King

Former Member
Colorado House of Representatives
Colorado Springs, Colorado

Kim Kozbial-Hess

Fourth-Grade Teacher
Hawkins Elementary School
Toledo, Ohio

Henry Kranendonk

Mathematics Curriculum Specialist
Milwaukee Public Schools
Milwaukee, Wisconsin

James S. Lanich

President
California Business for Education Excellence
Sacramento, California

Honorable Cynthia L. Nava

Senator
New Mexico State Senate
Las Cruces, New Mexico

Honorable Steven L. Paine

State Superintendent of Schools
West Virginia Department of Education
Charleston, West Virginia

Susan Pimentel

Educational Consultant
Hanover, New Hampshire

Andrew C. Porter

Dean
Graduate School of Education
University of Pennsylvania
Philadelphia, Pennsylvania

Honorable Kathleen Sebelius

Governor of Kansas
Topeka, Kansas

Warren T. Smith, Sr.

Vice President
Washington State Board of Education
Olympia, Washington

Mary Frances Taymans, SND

Executive Director
Secondary Schools Department
National Catholic Educational Association
Washington, D.C.

Oscar A. Troncoso

Principal
Anthony High School
Anthony, Texas

Grover J. Whitehurst (Ex officio)

Director
Institute of Education Sciences
U.S. Department of Education
Washington, D.C.

Charles E. Smith

Executive Director
National Assessment Governing Board
Washington, D.C.