

4

Sample Assessment Questions and Student Responses

This chapter presents sample questions and examples of student responses from the NAEP 2002 reading assessment. The complete reading passages to which the sample questions refer are provided in appendix D. Four representative questions, including both multiple-choice and constructed-response questions, are provided for each grade. For each question, both the framework-guided reading context and aspect are given. In the case of multiple-choice questions, the oval corresponding to the correct answer is filled in. Answers to constructed-response questions are accompanied by both a summary of the scoring criteria used to determine their rating and their actual assigned ratings. The student responses presented in this section were selected to illustrate how questions were scored. Additional passages and questions, as well as student performance data, detailed scoring guides, and sample student responses from previous NAEP assessments are available on the NAEP web site (<http://nces.ed.gov/nationsreportcard/itmrls>).

To indicate how students performed on the sample questions, each question included in this chapter is accompanied by a table presenting two types of performance data: (a) the overall percentage of students who answered successfully, and (b) the percentage of students who answered successfully within specific score ranges on the NAEP reading scale. The score ranges correspond to the three achievement level intervals—*Basic*, *Proficient*, and *Advanced*—as well as the range below *Basic*.

The sample questions are also marked on the item maps at the end of the chapter. The item map location of each multiple choice question identifies the scale score at which at least 74 percent of the students answered the question correctly. The item map location of each constructed-response question indicates the scale score at which at least 65 percent of the students reached a particular rating level.

Grade 4 Sample Assessment Questions and Results

Sample questions from the fourth-grade reading assessment include two multiple-choice, one short constructed-response, and one extended constructed-response question.

Information about the context and aspect of reading for each question shows how the item fits into the framework.

The fourth-grade reading comprehension questions presented here were based on the short story, “The Box in the Barn,” by Barbara Eckfield Connor. Jason, the story’s main character, learns a lesson about the risks of snooping when he accidentally lets loose a puppy he believes to be his sister’s birthday present. After a day of worry and guilt, Jason is relieved and excited to learn that his father has rescued the puppy, which turns out to be a surprise gift for the boy.

Grade 4 Sample question 1 (multiple-choice)

In sample question 1, students were asked to choose an answer that explains the character’s motivation. This item was easy for the students, with 77 percent of fourth-graders choosing the correct answer. This question appears on the item map at scale score 208.

When Megan spoke to Jason in the tall weeds, she was concerned that

- A she wouldn’t get enough presents
- B her dad wouldn’t get back in time for the party
- C something was wrong with Jason
- D the puppy was missing from the box

Reading Context:
Reading for Literary Experience

Reading Aspect:
Developing Interpretation

Table 4.1 Percentage scored correct for multiple-choice sample question 1, by achievement level range, grade 4: 2002

Grade 4	Percentage correct				
	Overall percentage correct	Below <i>Basic</i> 207 or below ¹	At <i>Basic</i> 208–237 ¹	At <i>Proficient</i> 238–267 ¹	At <i>Advanced</i> 268 or above ¹
	77	48	87	96	99

¹ NAEP reading composite scale range.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample question 2 (multiple-choice)

In sample question 2, students were asked to identify dialogue that illustrates a character's feelings within the story. Sixty percent of fourth-graders answered this question correctly. This question appears on the item map at scale score 241.

What does Megan say in the story that shows how she felt about Jason's getting a gift on her birthday?

- Ⓐ "Jason, Jason, I'm six years old."
- Ⓑ "Are you ok?"
- Ⓒ "Let's see what Dad wants."
- Ⓓ "Isn't he wonderful, Jason?"

Reading Context:
Reading for Literary Experience

Reading Aspect:
Examining Content and Structure

Table 4.2 Percentage scored correct for multiple-choice sample question 2, by achievement level range, grade 4: 2002

Overall percentage correct	Percentage correct			
	Below <i>Basic</i> 207 or below ¹	At <i>Basic</i> 208–237 ¹	At <i>Proficient</i> 238–267 ¹	At <i>Advanced</i> 268 or above ¹
60	37	63	80	90

¹ NAEP reading composite scale range.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample question 3 (short constructed-response)

This sample question asked students to demonstrate understanding of the story by predicting how one character might respond to a hypothetical situation. Responses to this question were scored as “Acceptable” or “Unacceptable.” Nearly two-thirds of fourth-graders’ responses were rated “Acceptable.” This question appears on the item map at scale score 220.

If the box had been empty when Jason opened it at the party, what would Jason most likely have said? Give examples from the story that support your answer.

Reading Context:

Reading for Literary Experience

Reading Aspect:

Examining Content and Structure

Table 4.3 Percentage scored “Acceptable” for short constructed-response sample question 3, by achievement level range, grade 4: 2002

Grade 4

Overall percentage “Acceptable”	Percentage “Acceptable”			
	Below <i>Basic</i> 207 or below ¹	At <i>Basic</i> 208–237 ¹	At <i>Proficient</i> 238–267 ¹	At <i>Advanced</i> 268 or above ¹
63	37	70	81	88

¹ NAEP reading composite scale range.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample “Acceptable” Response

Responses scored “Acceptable” gave story-related evidence to support the student’s reasoning. In this sample answer, the student notes that Jason seemed to be an honest boy.

If the box had been empty when Jason opened it at the party, what would Jason most likely have said? Give examples from the story that support your answer.

He would have said “I went to the back to find something to do, and I heard a strange noise, and went in to the barn to see what it was.” he sounded like a very honest boy.

Sample question 4 (extended constructed-response)

Sample question 4 assessed students' ability to understand character development by recognizing the different feelings presented in the story and the causes of those feelings. Answers to this question were scored with a four-level rating as "Extensive," "Essential," "Partial," or "Unsatisfactory." Students found this question somewhat difficult, with only 48 percent of fourth-graders scoring "Essential" or better. An "Essential" or better response to this item maps at the scale score 245.

From when Jason got up in the morning until he went to bed that night, his feelings changed as different things happened. Describe three different feelings that Jason had and explain what made him have those feelings.

Reading Context:
Reading for Literary Experience

Reading Aspect:
Developing Interpretation

Table 4.4a Percentage scored “Essential” or better for extended constructed-response sample question 4, by achievement level range, grade 4: 2002

Grade 4	Percentage “Essential” or better				
	Overall percentage “Essential” or better	Below <i>Basic</i> 207 or below ¹	At <i>Basic</i> 208–237 ¹	At <i>Proficient</i> 238–267 ¹	At <i>Advanced</i> 268 or above ¹
	48	17	46	70	88

¹ NAEP reading composite scale range.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample “Essential” Response

The following response is rated “Essential” because it identifies different feelings Jason experienced in response to changing events over the course of the day.

From when Jason got up in the morning until he went to bed that night, his feelings changed as different things happened. Describe three different feelings that Jason had and explain what made him have those feelings.

When he woke up he was in a happy mood. When the party came he was scared that the puppy would not be found and at the end of the day he was happy again.

Table 4.4b Percentage scored “Extensive” for extended constructed-response sample question 4, by achievement level range, grade 4: 2002

Overall percentage “Extensive”	Percentage “Extensive”			
	Below <i>Basic</i> 207 or below ¹	At <i>Basic</i> 208–237 ¹	At <i>Proficient</i> 238–267 ¹	At <i>Advanced</i> 268 or above ¹
1	#	#	1	4

Percentage rounds to 0.
¹ NAEP reading composite scale range.
 SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample “Extensive” Response

The following sample response is rated “Extensive” because it not only discusses three different feelings Jason had during the day, but also explains causes for each particular feeling, thereby demonstrating an in-depth understanding of Jason’s character.

From when Jason got up in the morning until he went to bed that night, his feelings changed as different things happened. Describe three different feelings that Jason had and explain what made him have those feelings.

He felt sad because he let the dog out and it ran away and he thought he ruined Megan's birthday.

He felt happy because his dad found the dog and because the dog was for him.

He felt wrong because he opened the box and his mother told him not to touch it.

Grade 8 Sample Assessment Questions and Results

Sample questions from the eighth-grade reading assessment include two multiple-choice questions, one short constructed-response question, and one extended constructed-response question.

These eighth-grade reading comprehension questions were based on “The Sharebots,” by Carl Zimmer. This article explains the work of a Brandeis University computer scientist, Maya Mataric, who programmed her “Nerd Herd,” a squad of 14 small robots, to socialize and cooperate for efficient task management.

Grade 8

Sample question 5 (multiple-choice)

Sample question 5 asked students to choose the statement of author’s purpose for the article. With an overall percentage correct of 82, this sample question was quite easy for the eighth-grade students taking the assessment. This question appears on the item map at scale score 243.

The main purpose of the article is to describe how robots can be programmed to

- Ⓐ locate metal pucks
- Ⓑ work with each other
- Ⓒ recharge their own batteries
- Ⓓ perform five basic behaviors

Reading Context:
Reading for Information

Reading Aspect:
Forming a General Understanding

Table 4.5 Percentage scored correct for multiple-choice sample question 5, by achievement level range, grade 8: 2002

Grade 8

Overall percentage correct	Percentage correct			
	Below <i>Basic</i> 242 or below ¹	At <i>Basic</i> 243–280 ¹	At <i>Proficient</i> 281–322 ¹	At <i>Advanced</i> 323 or above ¹
82	62	83	94	97

¹ NAEP reading composite scale range.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample question 6 (multiple-choice)

This sample question is a vocabulary item asking students to use contextual clues to determine the meaning of a word. Students taking the assessment found this item of average difficulty, with 57 percent of them answering this question correctly. This question appears on the item map at scale score 303.

The following sentence appears in the next-to-last paragraph of the article:

“With this simple social contract, the robots needed only 15 minutes of practice to become altruistic.”

Based on how the word is used in the article, which of the following best describes what it means to be altruistic?

- Ⓐ To engage in an experiment
- Ⓑ To provide assistance to others
- Ⓒ To work without taking frequent breaks
- Ⓓ To compete with others for the highest score

Reading Context:
Reading for Information

Reading Aspect:
Developing Interpretation

Table 4.6 Percentage scored correct for multiple-choice sample question 6, by achievement level range, grade 8: 2002

Overall percentage correct	Percentage correct			
	Below <i>Basic</i> 242 or below ¹	At <i>Basic</i> 243–280 ¹	At <i>Proficient</i> 281–322 ¹	At <i>Advanced</i> 323 or above ¹
57	41	51	73	91

¹ NAEP reading composite scale range.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample question 7 (short constructed-response)

Sample question 7 measures students' ability to judge the appropriateness of the article's title and to provide information from the text to support their reasoning. Answers to this question were scored with a three-level rating: evidence of "Full Comprehension," evidence of "Partial or Surface Comprehension," or evidence of "Little or No Comprehension." Students found this item difficult, with only 40 percent of the answers scoring at the level of "Full Comprehension." This question appears on the item map at scale score 310.

Do you think "The Sharebots" is a good title for this article?
Explain why or why not, using information from the article.

Reading Context:
Reading for Information

Reading Aspect:
Forming a General Understanding

Table 4.7 Percentage scored “Full Comprehension” for short constructed-response sample question 7, by achievement level range, grade 8: 2002

Grade 8		Percentage “Full Comprehension”			
Overall percentage “Full Comprehension”	Below <i>Basic</i> 242 or below ¹	At <i>Basic</i> 243–280 ¹	At <i>Proficient</i> 281–322 ¹	At <i>Advanced</i> 323 or above ¹	
40	16	37	60	82	

¹ NAEP reading composite scale range.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample “Full Comprehension” Response

The following sample response reflects “Full Comprehension” because it offers appropriate evidence from the article directly supporting the idea that the robots shared information.

Do you think “The Sharebots” is a good title for this article? Explain why or why not, using information from the article.

It is a good title because the robots share information on location of pucks and who retrieves them.

Sample question 8 (extended constructed-response)

This sample question required students to connect information from the text with their own background knowledge in order to compare and contrast the collaborative efforts of humans and sharebots. Responses to this item were scored with a four-level rating: "Extensive," "Essential," "Partial," or "Unsatisfactory." About half of the eighth-graders assessed provided responses rated as "Essential" or better. The "Extensive" response to this question appears on the item map at scale score 400.

Describe the similarities and differences between the way people work together and the way sharebots work together. Use examples from the article and from your own experiences in your description.

Reading Context:
Reading for Information

Reading Aspect:
Making Reader/Text Connections

Table 4.8a Percentage scored “Essential” or better for extended constructed-response sample question 8, by achievement level range, grade 8: 2002

Grade 8	Percentage “Essential” or better				
	Overall percentage “Essential” or better	Below <i>Basic</i> 242 or below ¹	At <i>Basic</i> 243–280 ¹	At <i>Proficient</i> 281–322 ¹	At <i>Advanced</i> 323 or above ¹
	51	21	49	72	90

¹ NAEP reading composite scale range.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample “Essential” Response

This sample answer is rated “Essential” because it uses information from the text to describe differences between sharebots and humans.

Describe the similarities and differences between the way people work together and the way sharebots work together. Use examples from the article and from your own experiences in your description.

The difference between the way people work together and sharebots work together is people are more civilized. Sharebots are machines, they can't think or reason, people can. That's the difference between people and sharebots.

Table 4.8b Percentage scored “Extensive” for extended constructed-response sample question 8, by achievement level range, grade 8: 2002

Grade 8	Percentage “Extensive”				
	Overall percentage “Extensive”	Below <i>Basic</i> 242 or below ¹	At <i>Basic</i> 243–280 ¹	At <i>Proficient</i> 281–322 ¹	At <i>Advanced</i> 323 or above ¹
	10	1	6	20	31

¹ NAEP reading composite scale range.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample “Extensive” Response

This sample answer is rated “Extensive” because it compares and contrasts humans and sharebots by offering information that goes beyond isolated behaviors.

Describe the similarities and differences between the way people work together and the way sharebots work together. Use examples from the article and from your own experiences in your description.

We all the same because when we have a set goal we work together to achieve it. This means we contact each other and help each other. This is what the sharebots are doing. We are different because we have minds and consciousness that tell us what to do. We aren't programmed for specific tasks. We also don't run on batteries.

Grade 12 Sample Assessment Questions and Results

Sample questions from the twelfth-grade reading assessment include one multiple-choice, two short constructed-response, and one extended constructed-response question.

The twelfth-grade reading comprehension questions presented here were based on

“Address to the Broadcasting Industry,” by Newton Minow. This selection is the text of Newton Minow’s 1961 speech to the National Association of Broadcasters, giving examples to support his indictment of American television programming as “a vast wasteland.”

Grade 12

Sample question 9 (multiple-choice)

In sample question 9, students were asked to choose the answer that best describes the kind of support Newton Minow used to defend his position. About three-quarters of the twelfth-graders assessed chose the correct answer for this item. This question appears on the item map as scale score 290.

Mr. Minow mainly supported his position with

- A personal opinions
- B rating statistics
- C recommendations from advertisers
- D newspaper articles

Reading Context:
Reading for Information

Reading Aspect:
Examining Content and Structure

Table 4.9 Percentage scored correct for multiple-choice sample question 9, by achievement level range, grade 12: 2002

Grade 12	Percentage correct				
	Overall percentage correct	Below <i>Basic</i> 264 or below ¹	At <i>Basic</i> 265–301 ¹	At <i>Proficient</i> 302–345 ¹	At <i>Advanced</i> 346 or above ¹
	72	52	71	84	92

¹ NAEP reading composite scale range.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample question 10 (short constructed-response)

Sample question 10 required students to link information across parts of the text to show their understanding of ways to resolve the problems in children’s programming. This item was scored with a three-level rating: evidence of “Full Comprehension,” evidence of “Partial or Surface Comprehension,” or evidence of “Little or No Comprehension.”

More than half of twelfth-graders provided responses that reflected “Full Comprehension.” This question appears on the item map at scale score 291.

According to Mr. Minow, how might the problems in children’s programming be solved?

Reading Context:
Reading for Information

Reading Aspect:
Developing Interpretation

Table 4.10 Percentage scored “Full Comprehension” for short constructed-response sample question 10, by achievement level range, grade 12: 2002

Grade 12				
Percentage “Full Comprehension”				
Overall percentage “Full Comprehension”	Below <i>Basic</i> 264 or below ¹	At <i>Basic</i> 265–301 ¹	At <i>Proficient</i> 302–345 ¹	At <i>Advanced</i> 346 or above ¹
61	27	60	82	96

¹ NAEP reading composite scale range.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample “Full Comprehension” Response

This sample answer is scored “Full Comprehension” because it demonstrates insight into the different problems affecting children’s programming and supplies at least one example from Minow’s speech.

According to Mr. Minow, how might the problems in children’s programming be solved?

If they took off the cartoons and violence and put on more educational programs to teach them about the world in a level they can understand.

Sample question 11 (short constructed-response)

This sample question measured students' ability to link information from across the text in order to explain Minow's meaning of "a vast wasteland." Answers to this question were scored with a three-level rating: evidence of "Full Comprehension," evidence of "Partial or Surface Comprehension," or evidence of "Little or No Comprehension." This was a difficult item for the students, with 27 percent earning "Full Comprehension." This question appears on the item map at scale score 336.

Why did Mr. Minow refer to television as "a vast wasteland"?
Give an example from the speech to support your answer.

Reading Context:
Reading for Information

Reading Aspect:
Developing Interpretation

Table 4.11 Percentage scored “Full Comprehension” for short constructed-response sample question 11, by achievement level range, grade 12: 2002

Grade 12		Percentage “Full Comprehension”			
Overall percentage “Full Comprehension”	Below <i>Basic</i> 264 or below ¹	At <i>Basic</i> 265–301 ¹	At <i>Proficient</i> 302–345 ¹	At <i>Advanced</i> 346 or above ¹	
27	5	22	43	63	

¹ NAEP reading composite scale range.
 SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample “Full Comprehension” Response

The following sample response is rated “Full Comprehension” because it demonstrates a clear understanding of Minow’s concern and provided a supporting example from the speech.

Why did Mr. Minow refer to television as “a vast wasteland”?
 Give an example from the speech to support your answer.

He called TV a vast wasteland because there is hardly anything worth watching. He used bad game shows and westerns as examples

Sample question 12 (extended constructed-response)

Sample question 12 asked students to use their own knowledge to judge the relevance of Minow’s critique of contemporary television programming. This question was scored with a four-level rating as “Extensive,” “Essential,” “Partial,” or “Unsatisfactory.” Students found this question fairly difficult, with 36 percent of their responses rated as “Essential” or higher. This question appears on the item map at scale score 387 for “Extensive” responses.

Imagine that Mr. Minow is preparing to deliver another address to the broadcasting industry. Would his original speech apply just as well to television programming today? Explain why or why not.

Reading Context:
Reading for Information

Reading Aspect:
Making Reader/Text Connections

Table 4.12a Percentage scored “Essential” or better for extended constructed-response sample question 12, by achievement level range, grade 12: 2002

Grade 12	Percentage “Essential” or better				
	Overall percentage “Essential” or better	Below <i>Basic</i> 264 or below ¹	At <i>Basic</i> 265–301 ¹	At <i>Proficient</i> 302–345 ¹	At <i>Advanced</i> 346 or above ¹
	36	10	29	56	79

¹ NAEP reading composite scale range.
 SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample “Essential” Response

This sample answer is rated “Essential” because it demonstrates a clear understanding of a major issue from the speech and generally relates that issue to present day television programming.

Imagine that Mr. Minow is preparing to deliver another address to the broadcasting industry. Would his original speech apply just as well to television programming today? Explain why or why not.

YES! TELEVISION HAS NOT CHANGED AS MUCH AS IT WAS A FEW YEARS AGO. THE SHOWS ARE MOSTLY THE SAME ALSO THE PUBLIC STILL HAS A VARIETY OF INTERESTS, AND TV STILL DOES NOT COVER THAT AND WE STILL HAVE A VARIETY OF JUNK SHOWS. SO YES THIS SPEECH WOULD APPLY MORE SO TODAY BECAUSE TELEVISION IS GETTING WORSE BY THE DAY.

Table 4.12b Percentage scored “Extensive” for extended constructed-response sample question 12, by achievement level range, grade 12: 2002

Grade 12

Overall percentage “Extensive”	Percentage “Extensive”			
	Below <i>Basic</i> 264 or below ¹	At <i>Basic</i> 265–301 ¹	At <i>Proficient</i> 302–345 ¹	At <i>Advanced</i> 346 or above ¹
10	1	6	17	40

¹ NAEP reading composite scale range.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Sample “Extensive” Response

This sample answer is rated “Extensive” because it demonstrates in-depth understanding of major issues from Minow’s speech and specifically relates those issues to present-day television programming.

Imagine that Mr. Minow is preparing to deliver another address to the broadcasting industry. Would his original speech apply just as well to television programming today? Explain why or why not.

I don't think this address would be as appropriate today as it originally was. Today's T.V. has alot more informative programs such as more news, history, wildlife etc. Today's programs are a little better than they used to be and there are more programs for young children to help them learn more about the world and school.

Maps of Selected Item Descriptions on the NAEP Reading Scale— Grades 4, 8, and 12

Item maps showing the description of particular items at the position along the NAEP reading composite scale where they are most likely to be successfully answered provide an illustration of the reading performance of fourth-, eighth- and twelfth-graders.¹ Descriptions of questions on the item map focus on the reading skills or abilities needed to answer the questions. For multiple-choice questions, the description indicates the comprehension demonstrated when students select the correct option. For constructed-response questions, the description indicates the degree of comprehension specified at different levels of the scoring criteria for that question. An examination of the descriptions may provide insight into the range of comprehension processes demonstrated by fourth-, eighth-, and twelfth-grade students.

For each question indicated on the map, students whose average scale scores fell at or above the scale point had a higher probability of successfully answering the question, while students whose average scale scores fell at or below that scale point had a lower probability of successfully answering that question. The map indicates the point at which individual comprehension questions were answered successfully by at least 65 percent of the students for constructed-response questions, or by at least 74 percent

of the students for multiple-choice questions.² For example, if a multiple-choice question, like the grade 4 sample question 1 on Table 4.1, maps at 208 on the scale, fourth-grade students with an average score of 208 or more have at least a 74 percent chance of answering this question correctly. In other words, out of every 100 students who scored at or above 208, at least 74 answered this question correctly. Although students scoring above the scale point have a higher probability of successfully answering the question, it does not mean that every student at or above 208 always answered this question correctly, nor does it mean that students below 208 always answered the question incorrectly. The item maps are useful indicators of higher or lower probability of successfully answering the question depending on students' overall ability as measured by the NAEP scale.

When considering information provided by item maps, it is important to be aware that the descriptions are based on comprehension questions that relate to specific reading passages. It is possible that questions intended to assess the same aspect of comprehension, when referring to different passages, would map at different points on the scale. In fact, one NAEP study found that even identically worded questions may be easier or harder when associated with different passages, suggesting that the difficulty of a question is related to its interaction with a particular passage.³

¹ For details on the procedures used to develop item maps, see Allen, N. R., Donoghue, J. R., and Schoeps, T. L. (1998). *The NAEP Technical Report*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

² The probability convention is set higher (at 74 percent) for multiple-choice questions to correct for the possibility of answering correctly by guessing.

³ Campell, J. R., and Donahue, P. L. (1997). *Students Selecting Stories: The Effects of Choice in Reading Assessment*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Figure 4.1 Map of selected item descriptions on the NAEP reading scale, grade 4: 2002



¹ Each grade 4 reading question in the 2002 reading assessment was mapped onto the NAEP 0–500 reading scale. The position of a question on the scale represents the average scale score attained by students who had a 65 percent probability of successfully answering a constructed-response question, or a 74 percent probability of correctly answering a four-option multiple-choice question. Only selected questions are presented. Scale score ranges for reading achievement levels are referenced on the map. For constructed-response questions, the question description represents students' performance at the scoring criteria level being mapped. NOTE: Regular type denotes a constructed-response question. Italic type denotes a multiple-choice question.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Figure 4.2 Map of selected item descriptions on the NAEP reading scale, grade 8: 2002

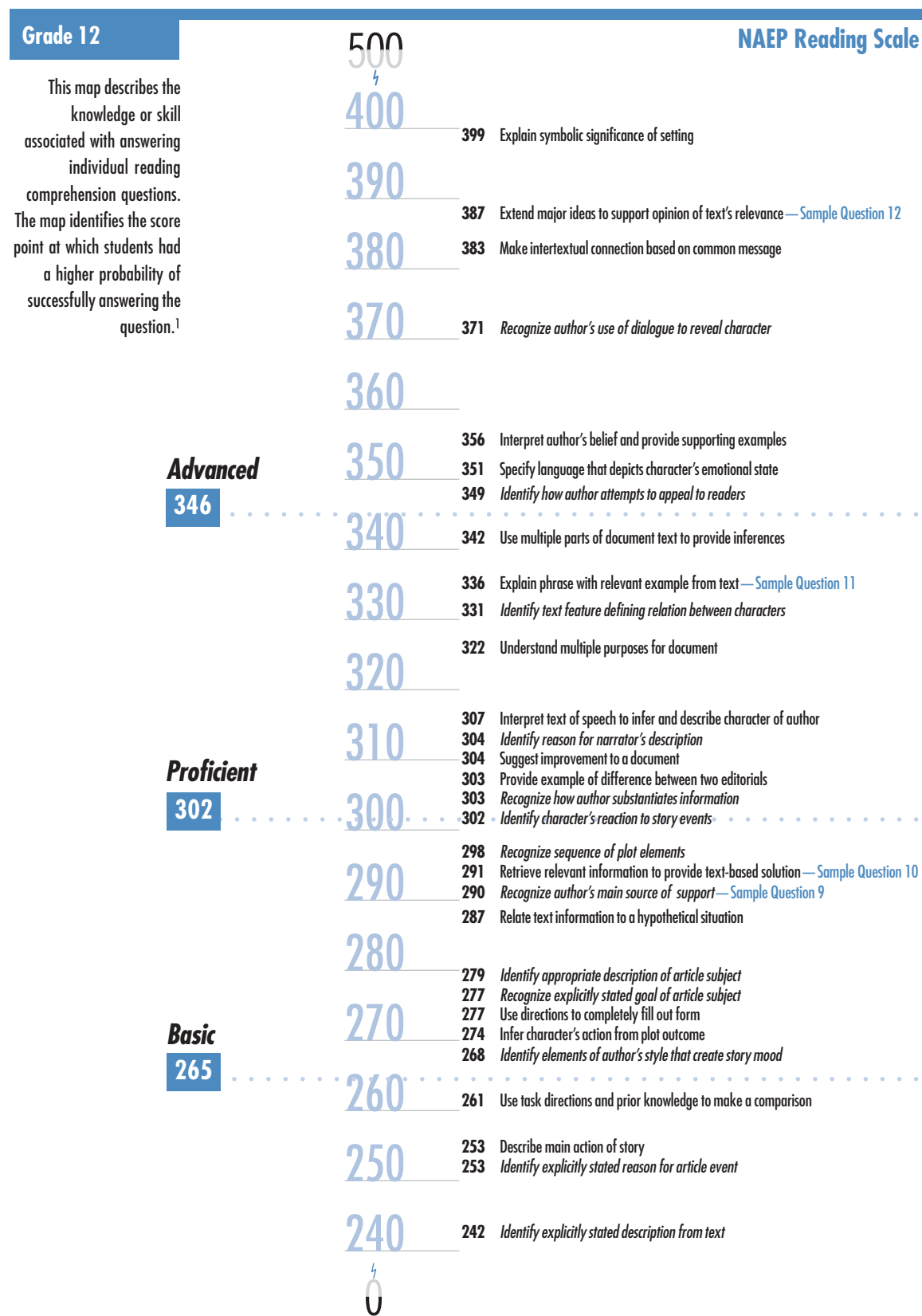


¹ Each grade 8 reading question in the 2002 reading assessment was mapped onto the NAEP 0–500 reading scale. The position of a question on the scale represents the average scale score attained by students who had a 65 percent probability of successfully answering a constructed-response question, or a 74 percent probability of correctly answering a four-option multiple-choice question. Only selected questions are presented. Scale score ranges for reading achievement levels are referenced on the map. For constructed-response questions, the question description represents students' performance at the scoring criteria level being mapped.

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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Figure 4.3 Map of selected item descriptions on the NAEP reading scale, grade 12: 2002



¹ Each grade 12 reading question in the 2002 reading assessment was mapped onto the NAEP 0–500 reading scale. The position of a question on the scale represents the average scale score attained by students who had a 65 percent probability of successfully answering a constructed-response question, or a 74 percent probability of correctly answering a four-option multiple-choice question. Only selected questions are presented. Scale score ranges for reading achievement levels are referenced on the map. For constructed-response questions, the question description represents students' performance at the scoring criteria level being mapped.

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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

A

Appendix A

Overview of Procedures Used for the NAEP 2002 Reading Assessment

This appendix provides an overview of the NAEP 2002 reading assessment's primary components—framework, development, administration, scoring, and analysis. A more extensive review of the procedures and methods used in the reading assessment will be included in the assessment procedures sections of the NAEP web site (<http://nces.ed.gov/nationsreportcard>).

The NAEP 2002 Reading Assessment

The National Assessment Governing Board (NAGB), created by Congress in 1988, is responsible for formulating policy for NAEP. NAGB is specifically charged with developing assessment objectives and test specifications. The design of the NAEP 2002 reading assessment follows the guidelines first provided in the framework developed for the 1992 assessment.¹ The framework underlying the 1992, 1994, 1998, 2000 (fourth grade only), and 2002 reading assessments reflects the expert opinions of educators and researchers about reading. Its purpose is to present an overview of the most essential outcomes of students' reading education. The development of this framework and the specifications that guided the development of the assessment 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 framework development

¹ National Assessment Governing Board. (2002). *Reading Framework for the 2003 National Assessment of Educational Progress*. Washington, DC: Author.

process was managed by the Council of Chief State School Officers (CCSSO) for NAGB.

The framework sets forth a broad definition of “reading literacy”—developing a general understanding of written text, thinking about text in different ways, and using a variety of text types for different purposes. In addition, the framework views reading as an interactive and constructive process involving the reader, the text, and the context of the reading experience. For example, readers may read stories to enjoy and appreciate the human experience, study science texts to form new hypotheses about knowledge, or use maps to gain information about specific places. NAEP reflects current definitions of literacy by differentiating among three contexts for reading and four aspects of reading. Contexts for reading and aspects of reading make up the foundation of the NAEP reading assessment.

The “contexts for reading” dimension of the NAEP reading framework provides guidance for the types of texts to be included in the assessment. Although many commonalities exist among the different reading texts, they do lead to real differences in what readers do. For example, when *reading for literary experience*, readers make complex, abstract summaries, and identify major themes. They describe the interactions of various literary elements (e.g., setting, plot, characters, and theme). When *reading for information*, readers critically judge the form and content of the text and explain their judgments. They also look for specific pieces of information. When *reading to perform a task*, readers search quickly for specific pieces of information.

The “aspects of reading” dimension of the NAEP reading framework provides guidance for the types of comprehension questions to be included in the assessment. The four aspects are 1) *forming a general understanding*, 2) *developing interpretation*, 3) *making reader/text connections*, and 4) *examining content and structure*. These four aspects represent different ways in which readers develop understanding of a text. In *forming a general understanding*, readers must consider the text as a whole and provide a global understanding of it. As readers engage in *developing interpretation*, they must extend initial impressions in order to develop a more complete understanding of what was read. This involves linking information across parts of a text or focusing on specific information. When *making reader/text connections*, the reader must connect information in the text with knowledge and experience. This might include applying ideas in the text to the real world. Finally, *examining content and structure* requires critically evaluating, comparing and contrasting, and understanding the effect of different text features and authorial devices.

Figure A.1 demonstrates the relationship between these reading contexts and aspects of reading in the NAEP reading assessment. Included in the figure are sample questions that illustrate how each aspect of reading is assessed within each reading context. (Note that reading to perform a task is not assessed at grade 4.)

Figure A.1 Sample NAEP questions, by aspects of reading and contexts for reading specified in the reading framework

Context for Reading	Aspect of Reading			
	Forming a general understanding	Developing interpretation	Making reader/text connections	Examining content and structure
Reading for literary experience	<i>What is the story/plot about?</i>	<i>How did this character change from the beginning to the end of the story?</i>	<i>What other character that you have read about had a similar problem?</i>	<i>What is the mood of this story and how does the author use language to achieve it?</i>
Reading for information	<i>What point is the author making about this topic?</i>	<i>What caused this change?</i>	<i>What other event in history or recent news is similar to this one?</i>	<i>Is this author biased? Support your answer with information about this article.</i>
Reading to perform a task	<i>What time can you get a nonstop flight to X?</i>	<i>What must you do before step 3?</i>	<i>Describe a situation in which you would omit step 5?</i>	<i>Is the information in this brochure easy to use?</i>

SOURCE: National Assessment Governing Board. (2002). *Reading Framework for the 2003 National Assessment of Educational Progress*. Washington, DC: Author.

The assessment framework specifies not only the particular dimensions of reading literacy to be measured, but also the percentage of assessment questions that should be devoted to each. The target percentage distribution for contexts of reading and aspects of reading as specified in the framework, along with the actual percentage distribution in the assessment, are presented in tables A.1 and A.2.

The actual content of the assessment has varied from the targeted distribution, with reading for literary experience falling below the target proportions and reading for information falling above the target proportions specified in the framework. The reading instrument development panel overseeing the development of the assessment recognized this variance but felt strongly that assessment questions must be sensitive to the unique elements of the authentic reading materials being used. Thus, the distribution of question classifications will vary across reading passages and reading purposes.

Table A.1 Target and actual percentage distribution of questions, by context for reading, grades 4, 8, and 12: 2002

		Context for Reading		
		Reading for literary experience	Reading for information	Reading to perform a task
Grade 4	Target	55	45	†
	Actual	50	50	†
Grade 8	Target	40	40	20
	Actual	27	43	30
Grade 12	Target	35	45	20
	Actual	24	49	27

† Reading to perform a task was not assessed at grade 4.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Table A.2 Target and actual percentage distribution of questions, by aspect of reading, grades 4, 8, and 12: 2002

		Aspect of Reading		
		Forming a general understanding/ Developing interpretation	Making reader/text connections	Examining content and structure
Grade 4	Target	60	15	25
	Actual	59	18	24
Grade 8	Target	55	15	30
	Actual	54	18	28
Grade 12	Target	50	15	35
	Actual	52	18	31

NOTE: Actual percentages are based on the classifications agreed upon by NAEP's Instrument Development Panel. It is recognized that making discrete classifications for these categories is difficult and that independent efforts to classify NAEP questions have led to different results.

Percentages may not add to 100, due to rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

The Assessment Design

Each student who participated in the reading assessment received a booklet containing three or four sections: a set of general background questions, a set of subject-specific background questions, and one or two sets of questions assessing students' comprehension of a text or texts. The sets of questions assessing students' comprehension are referred to as "blocks." Each block contains one or more reading passages and a set of comprehension questions. At grades 8 and 12, students were given either two 25-minute blocks or one 50-minute block. At grade 4, however, only 25-minute blocks were used.

The blocks contain a combination of multiple-choice and constructed-response questions. Multiple-choice questions require students to select the best answer from a set of four options. Constructed-response questions require students to provide their own written response to an open-ended question. Short constructed-response questions may require a response of only a sentence or two for the answer to be considered complete. Extended constructed-response questions, however, may require a response of a paragraph or more for the answer to receive full credit. Each constructed-response question has its own unique scoring guide that is used by trained scorers to rate students' responses. (See the "Data Collection and Scoring" section of this appendix.)

The grade 4 assessment consisted of eight 25-minute blocks: four blocks of "literary" texts and questions and four blocks of "informative" texts and questions. Each block contained at least one passage corresponding to one of the contexts for reading and 9–12 multiple-

choice and constructed-response questions. In each block, one of the constructed-response questions required an extended response. As a whole, the 2002 fourth-grade assessment consisted of 49 multiple-choice questions, 45 short constructed-response questions, and 8 extended constructed-response questions.

The grade 8 assessment consisted of nine 25-minute blocks (three literary, three informative, and three task) and one 50-minute block (informative). Each block contained at least one passage corresponding to one of the contexts for reading and 8 to 13 multiple-choice and constructed-response questions. Each block contained at least one extended constructed-response question. As a whole, the eighth-grade assessment consisted of 58 multiple-choice questions, 68 short constructed-response questions, and 15 extended constructed-response questions.

The grade 12 assessment consisted of nine 25-minute blocks (three literary, three informative, and three task) and two 50-minute blocks (informative). The blocks contained at least one passage and 8 to 16 multiple-choice and constructed-response questions. Each block contained at least one extended constructed-response question. As a whole, the twelfth-grade assessment contained 40 multiple-choice questions, 61 short constructed-response questions, and 13 extended constructed-response questions.

The assessment design allowed maximum coverage of reading abilities at each grade, while minimizing the time burden for any one student. This was accomplished through the use of matrix sampling of items in which representative samples of students took various portions of the entire

pool of assessment questions. Individual students are required to take only a small portion, but the aggregate results across the entire assessment allow for broad reporting of reading 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 receive different blocks of passages and comprehension questions in their booklets according to a procedure called “partially balanced incomplete block (PBIB) spiraling.” This 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 context for reading. Blocks were balanced within each context for reading and were partially balanced across contexts for reading. 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 booklet.

In addition to the student assessment booklets, three other instruments provided data relating to the assessment—a teacher questionnaire, a school questionnaire, and a questionnaire for students with disabilities and limited English proficient students (SD/LEP). The teacher questionnaire was administered to teachers of fourth- and eighth-grade students participating in the assessment and included four sections. The first section focused on teacher’s background; the second section on instruction; the third section on professional development; and the fourth section on standards and assessment.

The school questionnaire was given to the principal or other administrator in each participating school and included questions related to school policies, programs, and the composition and background of the student body.

The SD/LEP questionnaire was completed by a school staff member knowledgeable about those students who were selected to participate in the assessment and who were identified as having an Individualized Education Program (IEP) or equivalent plan, or being limited English proficient (LEP). An SD/LEP questionnaire was completed for each identified student regardless of whether the student participated in the assessment. Each SD/LEP questionnaire asked about the student and the special programs in which he or she participated.

NAEP Samples

National Sample

The national results presented in this report are based on nationally representative probability samples of fourth-, eighth-, and twelfth-grade students. At grades 4 and 8, the national sample in 2002 was a subset of the combined sample of students assessed in each participating state, plus an additional sample from the states that did not participate in the state assessment as well as a private school sample. This represents a change from previous assessments in which the national and state samples were independent. At grade 12, the sample was chosen using a stratified two-stage design that involved sampling students from selected schools (public and nonpublic) across the country.

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 between the student samples and the respective populations from which they were drawn. Sampling weights account for disproportionate representation due to the oversampling of students who attend schools with high concentrations of Black and/or Hispanic students and students who attend nonpublic schools. Among other uses, sampling weights also account for lower sampling rates for very small schools and are used to adjust for school and student nonresponse.²

Unlike the 1998 and 2000 national assessments, which featured the collection of data from samples of students where assessment accommodations for special-needs students were not permitted and from samples of students where accommodations for special-needs students were permitted, the 2002 national assessment has only samples of students where accommodations were permitted. NAEP inclusion rules were applied, and accommodations were offered when a student had an Individualized Education Program (IEP) because of a disability, was protected under

Section 504 of the Rehabilitation Act of 1973³ because of disability and/or was identified as being a limited English proficient student (LEP); all other students were asked to participate in the assessment under standard conditions. Prior to 1998, testing accommodations (e.g., extended time, small group testing) were not permitted for special-needs students selected to participate in the NAEP reading assessments.

Table A.3 shows the number of students included in the national samples for the NAEP reading assessments at each grade level. The 2002 reading assessment has only the sample of students in which accommodations were permitted. For the 1998 and 2000 assessments, the table includes the number of students in the sample in which accommodations were not permitted and the number of students in the sample in which accommodations were permitted. The table shows that the same non-SD and/or non-LEP students were included in both samples; only the SD and/or LEP students differed between the two samples. The 1992 and 1994 design differed from more recent assessment years in that the SD and/or LEP students were assessed in standard conditions and accommodations were not permitted.

² Additional details regarding the design and structure of the national and state samples will be included in the technical documentation section of the NAEP web site at <http://nces.ed.gov/nationsreportcard>.

³ 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.

Table A.3 Number of students assessed, by sample type, special needs status and accommodation option, grades 4, 8, and 12 public and non public schools: 1992–2002

	1992	1994	1998		2000		2002
	Accommodations not permitted sample	Accommodations not permitted sample	Accommodations not permitted sample	Accommodations permitted sample	Accommodations not permitted sample	Accommodations permitted sample	Accommodations permitted sample
Grade 4							
Total students assessed	6,314	7,382	7,672	7,812	7,914	8,074	140,487
Non-SD/LEP ¹ students assessed	6,051	6,783	7,232		7,484		122,721
SD/LEP ¹ students assessed without accommodations	263	599	440	413	430	476	11,913
SD/LEP ¹ students assessed with accommodations	†	†	†	167	†	114	5,853
Grade 8							
Total students assessed	9,464	10,135	11,051	11,193	—	—	115,176
Non-SD/LEP ¹ students assessed	9,184	9,676	10,309		—	—	102,174
SD/LEP ¹ students assessed without accommodations	280	459	742	678	—	—	8,598
SD/LEP ¹ students assessed with accommodations	†	†	†	206	—	—	4,404
Grade 12							
Total students assessed	9,856	9,935	12,675	12,760	—	—	14,724
Non-SD/LEP ¹ students assessed	9,726	9,646	12,112		—	—	13,784
SD/LEP ¹ students assessed without accommodations	130	289	563	532	—	—	673
SD/LEP ¹ students assessed with accommodations	†	†	†	116	—	—	267

— Data were not collected at grades 8 and 12 in 2000.

† Accommodations were not permitted in this sample.

¹ Students with disabilities/limited English proficient students.

NOTE: The sample sizes at grades 4 and 8 are larger in 2002 than in previous years because the 2002 national sample was based on the combined sample of students assessed in each participating state, plus an additional sample from non-participating states as well as a sample of private schools.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, 2000, and 2002 Reading Assessments.

Table A.4 provides a summary of the 2002 national school and student participation rates for the reading assessment sample. Participation rates are presented for public and nonpublic schools both individually and combined. The first rate is the weighted percentage of schools participating in the assessment before substitution of demographically similar schools.⁴ This rate is based only on the number of schools that were initially selected for the assessment. The numerator of this rate is the sum of the number of students represented by each initially selected school that participated in the assessment. The denominator is the sum of the number of students represented by each of the initially selected schools that had grade-eligible students enrolled.

The second school participation rate is the weighted participation rate after substitution. The numerator of this rate is the sum of the number of students represented by each of the participating schools, whether originally selected or selected as a substitute for a school that chose not to participate. The denominator is the sum of the estimated number of students represented by each of the initially selected schools that had eligible students enrolled (this is the same as that for the weighted participation rate for the sample of schools before substitution). The denominator for these two rates is an estimate of the number of students eligible for the assess-

ment, from all schools in the nation with eligible students enrolled. Because of the common denominators, the weighted participation rate after substitution is at least as great as the weighted participation rate before substitution.

Also presented in table A.4 are weighted student participation rates. The numerator of this rate is the sum of the number of students that each student represents (across all students assessed in either an initial session or a makeup session). The denominator of this rate is the sum of the number of students represented in the sample, across all eligible sampled students in participating schools. The overall participation rates take into account the weighted percentage of school participation before or after substitution and the weighted percentage of student participation after makeup sessions.

For the grade 12 national sample, where school and student response rates did not meet NCES standards, an extensive analysis was conducted that examined, among other factors, the potential for nonresponse bias at both the school and student level. No evidence of any significant potential for either school or student nonresponse bias was found. Results of these analyses, as well as nonresponse bias analyses for the grades 4 and 8 national samples will be included in the technical documentation.

⁴ The initial base sampling weights were used in weighting the percentages of participating schools and students. An attempt was made to preselect (before field processes began) a maximum of two substitute schools for each sampled public school (one in-district and one out-of-district) and each sampled Catholic school, and one for each sampled nonpublic school other than Catholic. To minimize bias, a substitute school resembled the original selection as much as possible in affiliation, estimated number of grade-eligible students, and minority composition.

Table A.4 National school and student participation rates, by type of school, grades 4, 8, and 12: 2002

	Weighted school participation			Student participation		Overall participation rate	
	Percentage before substitution	Percentage after substitution	Number of schools participating after substitution	Weighted percentage student participation	Number of students assessed	Before substitution	After substitution
Grade 4							
Combined national	84	85	5,518	94	140,487	79	80
Public	85	85	5,067	94	133,805	80	80
Nonpublic	74	81	451	95	5,578	71	77
Grade 8							
Combined national	82	83	4,706	92	115,176	75	76
Public	83	84	4,208	91	109,356	76	77
Nonpublic	68	76	498	95	5,320	65	72
Grade 12							
Combined national	74	75	725	74	14,724	55	55
Public	76	76	443	72	9,204	55	55
Nonpublic	55	59	282	88	5,520	48	52

NOTE: The number of students in the combined national total at grades 4 and 8 includes students in the Department of Defense domestic schools located within the U.S. and Bureau of Indian Affairs schools that are not included as part of either the public or nonpublic totals.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

State Samples

The results provided in this report of the 2002 state assessment in reading are based on state-level samples of fourth- and eighth-grade public-school students. The samples were selected using a two-stage sample design that first selected schools within participating states and other jurisdictions and then 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. Tables A.5 and A.6 contain the unweighted number of participating schools and students, as well as weighted school and student participation rates for the state samples at grades 4 and 8 respectively.

District Samples

Results from the 2002 reading assessments will also be reported (on a trial basis) in a forthcoming report on district-level samples of fourth- and eighth-grade students in the large urban school districts that participated in the Trial Urban District Assessment (Atlanta, Chicago, Houston, Los Angeles, and New York City). The sample of students in the urban school districts represents an augmentation to the sample of students who would “normally” be selected as part of state samples. These samples allow reliable subgroup reporting in these districts. Furthermore, all students at “lower” sampling levels are assumed to be part of “higher-level” samples. For example, Houston is one of the urban districts included in the Trial Urban District Assessment. Data from students tested in the Houston sample were used to report results for Houston, but also contributed to the Texas and national estimates.

Table A.5 School and student participation rates, grade 4 public schools: By state, 2002

Grade 4	Weighted school participation			Student participation		Overall participation rate	
	Percentage before substitution	Percentage after substitution	Number of schools participating after substitution	Weighted percentage student participation	Number of students assessed	Before substitution	After substitution
Nation (Public)	85	85	5,067	94	133,805	80	80
Alabama	84	96	108	95	3,684	80	92
Arizona	91	91	105	91	3,105	83	83
Arkansas	99	99	107	94	2,779	93	93
California †	72	72	143	95	4,016	68	68
Connecticut	100	100	108	95	3,266	95	95
Delaware	100	100	86	94	3,895	94	94
Florida	100	100	103	95	3,226	95	95
Georgia	100	100	152	95	4,919	95	95
Hawaii	100	100	111	96	3,603	96	96
Idaho	87	87	98	95	2,710	82	82
Illinois †	57	57	117	93	3,117	53	53
Indiana	99	99	112	94	3,469	93	93
Iowa †	77	77	86	95	1,930	73	73
Kansas †	73	73	84	96	1,938	70	70
Kentucky	96	96	106	96	3,262	92	92
Louisiana	99	99	116	96	3,116	95	95
Maine	88	88	98	94	1,964	83	83
Maryland	100	100	105	93	2,844	93	93
Massachusetts	100	100	111	95	3,236	95	95
Michigan	98	99	110	92	2,974	90	91
Minnesota †	77	77	84	95	2,598	73	74
Mississippi	95	95	104	95	3,091	90	90
Missouri	94	100	113	94	2,973	89	94
Montana †	75	75	79	95	1,342	71	71
Nebraska	95	95	91	96	1,540	91	91
Nevada	100	100	114	93	3,447	93	93
New Mexico	93	93	104	94	2,316	87	87
New York †	77	77	90	91	2,401	70	70
North Carolina	100	100	112	94	3,276	94	94
North Dakota †	82	82	164	96	2,422	79	79
Ohio	95	95	107	93	2,722	89	89
Oklahoma	99	99	132	95	3,352	94	94
Oregon	85	88	100	94	2,675	80	83
Pennsylvania	100	100	114	94	3,383	94	94
Rhode Island	100	100	113	94	3,551	94	94
South Carolina	99	99	105	95	2,473	94	94
Tennessee †	78	78	92	96	3,022	75	75
Texas	89	89	139	95	3,637	84	84
Utah	100	100	111	94	3,652	94	94
Vermont	90	90	106	95	1,690	85	85
Virginia	100	100	109	95	3,029	95	95
Washington †	75	75	85	95	2,444	71	71
West Virginia	99	99	136	96	2,348	95	95
Wisconsin †	55	55	63	95	1,475	52	52
Wyoming	100	100	162	95	2,786	95	95
Other Jurisdictions							
District of Columbia	100	100	117	90	2,554	90	90
DDESS ¹	99	99	39	96	1,351	95	95
DoDDS ²	99	99	91	95	2,924	94	94
Guam	100	100	25	96	1,216	96	96
Virgin Islands	100	100	24	95	738	95	95

† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2002.

¹ Department of Defense Domestic Dependent Elementary and Secondary Schools.

² Department of Defense Dependents Schools (Overseas).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Table A.6 School and student participation rates, grade 8 public schools: By state, 2002

Grade 8	Weighted school participation			Student participation		Overall participation rate	
	Percentage before substitution	Percentage after substitution	Number of schools participating after substitution	Weighted percentage student participation	Number of students assessed	Before substitution	After substitution
Nation (Public)	83	84	4,208	91	109,356	76	77
Alabama	80	93	100	93	2,602	75	87
Arizona	93	93	110	88	2,451	82	82
Arkansas	99	99	103	91	2,454	90	90
California †	71	71	125	90	3,124	64	64
Connecticut	100	100	104	92	2,682	92	92
Delaware	100	100	35	90	3,850	90	90
Florida	100	100	105	91	2,633	91	91
Georgia	100	100	111	93	3,756	93	93
Hawaii	100	100	55	93	2,656	93	93
Idaho	86	86	80	93	2,390	80	80
Illinois †	56	56	106	90	2,373	51	51
Indiana	98	98	101	91	2,535	89	89
Kansas †	72	72	83	93	1,827	67	67
Kentucky	96	96	100	94	2,461	90	90
Louisiana	98	98	98	93	2,252	91	91
Maine	94	94	101	92	2,522	86	86
Maryland	93	93	99	90	2,451	84	84
Massachusetts	98	98	104	93	2,576	91	91
Michigan	98	98	104	88	2,383	86	86
Minnesota †	66	66	67	91	1,657	60	60
Mississippi	94	94	96	93	2,415	87	87
Missouri	92	96	114	91	2,481	84	88
Montana †	76	76	73	94	1,849	71	71
Nebraska	99	99	103	92	2,139	91	91
Nevada	100	100	64	88	2,536	88	88
New Mexico	93	93	91	92	2,265	86	86
New York †	71	71	84	88	1,867	63	63
North Carolina	100	100	106	93	2,540	93	93
North Dakota †	77	77	110	94	1,949	73	73
Ohio	96	96	94	90	2,319	87	87
Oklahoma	100	100	123	92	2,493	92	92
Oregon †	78	78	85	91	1,918	71	71
Pennsylvania	100	100	104	92	2,720	92	92
Rhode Island	100	100	55	89	2,552	89	89
South Carolina	97	97	99	93	2,189	90	90
Tennessee †	74	74	82	92	2,047	69	69
Texas	92	92	127	93	3,258	85	85
Utah	100	100	93	92	2,683	92	92
Vermont	91	91	99	92	2,378	84	84
Virginia	100	100	103	92	2,546	92	92
Washington †	74	74	80	90	1,897	66	66
West Virginia	92	92	97	92	2,166	85	85
Wisconsin †	66	66	75	92	1,718	61	61
Wyoming	100	100	78	92	2,579	92	92
Other Jurisdictions							
American Samoa	100	100	22	96	460	96	96
District of Columbia	100	100	36	85	1,638	85	85
DDESS ¹	99	99	14	96	701	94	94
DoDDS ²	99	99	55	95	2,090	94	94
Guam	100	100	7	94	1,011	94	94
Virgin Islands	100	100	8	93	567	93	93

† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2002.

¹ Department of Defense Domestic Dependent Elementary and Secondary Schools.

² Department of Defense Dependents Schools (Overseas).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Standards for State Sample Participation and Reporting of Results

In carrying out the 2002 state assessment program, the National Center for Education Statistics (NCES) established participation rate standards that jurisdictions were required to meet in order for their results to be reported. NCES also established addi-

tional standards that required the annotation of published results for jurisdictions whose sample participation rates were low enough to raise concerns about their representativeness. The NCES guidelines used to report results in the state assessments, and the guidelines for notation when there is some risk of nonresponse bias in the reported results, are presented in this section.

Guideline 1

The publication of NAEP results

The conditions that will result in the publication of a jurisdiction's results are presented below.

Guideline 1 - Publication of Public School Results

A jurisdiction will have its public school results published in the 2002 NAEP reading report card (or in other reports that include all state-level results) if and only if its weighted participation rate for the initial sample of public schools is greater than or equal to 70 percent. Similarly, a jurisdiction will receive a separate NAEP State Report if and only if its weighted participation rate for the initial sample of public schools is greater than or equal to 70 percent.

Discussion: If a jurisdiction's public school participation rate for the initial sample of schools is below 70 percent, there is a substantial possibility that bias will be introduced into the assessment results. This possibility remains even after making statistical adjustments to compensate for school nonparticipation. There remains the likelihood that, in aggregate, the substitute schools are sufficiently dissimilar from the originals they are replacing and represent too great a proportion of the population to discount such a difference. Similarly, the assumptions underlying the use of statistical adjustments to compensate for nonparticipation are likely to be significantly violated if the initial response rate falls below the 70 percent level. Guideline 1 takes this into consideration. This guideline is congruent with current NAGB policy, which requires that data for jurisdictions that do not have a 70 percent before-substitution participation rate be reported "in a different format," and with the Education Information Advisory Committee (EIAC) resolution, which calls for data from such jurisdictions not to be published.

The following guidelines concerning school and student participation rates in the NAEP state assessment program were established to address four significant ways in which nonresponse bias could be introduced into the jurisdiction sample estimates. The four significant ways include overall school nonresponse, strata-specific school nonresponse, overall student

nonresponse and strata-specific student nonresponse. Presented on the following pages are the conditions that will result in a jurisdiction's receiving a notation in the 2002 reports. Note that in order for a jurisdiction's results to be published with no notations, that jurisdiction must satisfy all guidelines.

Guideline 2

Reporting school and student participation rates with possible bias due to school nonresponse

Guideline 2 - Notation for Overall Public School Participation Rate

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

Discussion: For jurisdictions that did not use substitute schools, the participation rates are based on participating schools from the original sample. In these situations, the NCES standards specify weighted school participation rates of at least 85 percent to guard against potential bias due to school nonresponse. Thus the first part of these guidelines, referring to the weighted school participation rate for the initial sample of schools, is in direct accordance with NCES standards.

To help ensure adequate sample representation for each jurisdiction participating in the NAEP 2002 state assessments, NAEP provided substitutes for nonparticipating public schools. For jurisdictions that used substitute schools, the assessment results will be based on the student data from all schools participating from both the original sample and the list of substitutes (unless both an initial school and its substitute eventually participated, in which case only the data from the initial school will be used).

The NCES standards do not explicitly address the use of substitute schools to replace initially selected schools that decide not to participate in the assessment. However, considerable technical consideration was given to this issue. Even though the characteristics of the substitute schools were matched as closely as possible to the characteristics of the initially selected schools, substitution does not entirely eliminate bias due to the nonparticipation of initially selected schools. Thus, for the weighted school participation rates including substitute schools, the guidelines were set at 90 percent.

If a jurisdiction meets either standard (i.e., 85 percent or higher prior to substitution or 90 percent or higher after substitution), there will be no notation for the relevant overall school participation rate.

Guideline 3

Important segments of the jurisdiction's student population that must be adequately represented to avoid possible nonresponse bias

Guideline 3 - Notation for Strata-Specific Public School Participation Rates

A jurisdiction that is not already receiving a notation under Guideline 2 will receive a notation if the sample of public schools included a class of schools with similar characteristics that had a weighted participation rate (after substitution) of below 80 percent, and from which the nonparticipating schools together accounted for more than 5 percent of the jurisdiction's total weighted sample of public schools. The classes of schools from each of which a jurisdiction needed minimum school participation levels were determined by degree of urbanization, minority enrollment, and median household income of the area in which the school is located.

Discussion: The NCES standards specify that attention should be given to the representativeness of the sample coverage. Thus, if some important segment of the jurisdiction's population is not adequately represented, it is of concern, regardless of the overall participation rate.

If nonparticipating schools are concentrated within a particular class of schools, the potential for substantial bias remains, even if the overall level of school participation appears to be satisfactory. Nonresponse adjustment cells for public schools have been formed within each jurisdiction, and the schools within each cell are similar with respect to minority enrollment, degree of urbanization, and/or median household income, as appropriate for each jurisdiction.

If the weighted response rate, after substitution, for a single adjustment cell falls below 80 percent, and more than 5 percent (weighted) of the sampled schools are nonparticipants from such a cell, the potential for nonresponse bias is too great. This guideline is based on the NCES standard for stratum-specific school response rates.

Guideline 4

Possible student nonresponse bias

Guideline 4 - Notation for Overall Student Participation Rate in Public Schools

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

Discussion: This guideline follows the NCES standard of 85 percent for overall student participation rates. The weighted student participation rate is based on all eligible students from initially selected or substitute schools who participated in the assessment in either an initial session or a make-up session. If the rate falls below 85 percent, the potential for bias due to students' nonresponse is too great.

Guideline 5

Possible nonresponse bias from inadequately represented strata

Guideline 5 - Notation for Strata-Specific Student Participation Rates in Public Schools

A jurisdiction that is not already receiving a notation under Guideline 4 will receive a notation if the sampled students within participating public schools included a class of students with similar characteristics that had a weighted student response rate of below 80 percent, and from which the nonresponding students together accounted for more than 5 percent of the jurisdiction's weighted assessable public school student sample. Student groups from which a jurisdiction needed minimum levels of participation were determined by the age of the student, whether or not the student was classified as a student with a disability (SD) or limited English proficient (LEP), and the type of assessment session, as well as school level of urbanization, minority enrollment, and median household income of the area in which the school is located.

Discussion: This guideline addresses the fact that if nonparticipating students are concentrated within a particular class of students, the potential for substantial bias remains, even if the overall student participation level appears to be satisfactory. Student nonresponse adjustment cells have been formed using the school-level nonresponse adjustment cells, together with the student's age and the nature of the assessment session.

If the weighted response rate for a single adjustment cell falls below 80 percent, and more than 5 percent (weighted) of the invited students who do not participate in the assessment are from such a cell, the potential for nonresponse bias is too great. This guideline is based on the NCES standard for stratum-specific student response rates.

At both the fourth and eighth grades, two states, Illinois and Wisconsin, did not meet the initial public-school participation rate of 70 percent. In addition, one state, Minnesota, did not meet this standard at the eighth grade. Results for these jurisdictions are not included with the findings reported for the state NAEP 2002 reading assessment.

Nine jurisdictions at grade 4 did not meet the second guideline for notation (i.e., the weighted participation rate for the

initial sample of schools was below 85 percent and the weighted school participation rate after substitution was below 90 percent): California, Iowa, Kansas, Minnesota, Montana, New York, North Dakota, Tennessee, and Washington. At grade 8, eight jurisdictions did not meet this guideline: California, Kansas, Montana, New York, North Dakota, Oregon, Tennessee, and Washington. Results for each of these jurisdictions at the appropriate grade level are shown with a notation indicating possible bias related to nonresponse.

Students with Disabilities (SD) and/or Limited English Proficient (LEP) Students

It is NAEP's intent 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 except in the following cases:

- the school's IEP team determined that the student could not participate;
- the student's cognitive functioning was so severely impaired that she or he could not participate;
- the student's IEP required that the student had to be tested with an accommodation or adaptation that NAEP does not allow and that the student could not demonstrate his or her knowledge without that accommodation.

All LEP students who received academic instruction in English for three years or more were to be included in the assessment. Those LEP students who received instruction in English for fewer than three years were to be included unless school staff judged them to be incapable of participating in the assessment in English.

Participation of SD and/or LEP 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 LEP students 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 in the last decade and gained momentum 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 in order to best show what they know and can do.⁵ 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 limited English proficient students in the population has increased, some states have started offering accom-

⁵ Office of Special Education Programs. (1997). *Nineteenth Annual Report to Congress on the Implementation of the Individuals With Disabilities Education Act*. Washington, DC: U. S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

modations, such as translations of assessments or the use of bilingual dictionaries as part of assessments.

Before 1996, NAEP did not allow any testing under nonstandard conditions (i.e., 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 grew more widespread, the failure to provide accommodations led to increasing levels of exclusion in the assessment. Such increases posed two threats to the program: 1) they threatened the stability of trend lines (because excluding more students in one year than the next might lead to apparent rather than real gains), and 2) they made NAEP samples less than optimally representative of target populations.

NAEP reacted to this challenge by adopting a multipart strategy. The program had to move toward allowing the same assessment accommodations that were afforded students in state and district testing programs in order for NAEP samples to be as inclusive as possible. However, allowing accommodations represents a change in testing conditions that may affect measurement of changes over time. Therefore, beginning with the 1996 national assessments and the 1998 state assessments 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. In addition to the samples where accommodations were not permitted, parallel samples in which accommodations were permitted were also assessed. By having two overlapping samples and two sets of related data points, NAEP could meet two core program goals.⁶ First, data trends could be maintained. Second, parallel trend lines could be set in ways that ensure that in future years the program will be able to use the most inclusive practices possible and mirror the procedures used by most state and district assessments. Beginning in 2002, NAEP uses only the more inclusive samples in which assessment accommodations are permitted.

In reading, national and state data from 1992, 1994, and 1998 are reported for the sample in which accommodations were not permitted. National and state data for the sample in which accommodations were permitted are reported for 1998 and 2002. National-only data at grade 4 for both accommodated and unaccommodated samples are reported for 2000.

In order 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

⁶ The two samples are described as “overlapping” because, in 1998 and 2000, the same group of non-SD and/or LEP students were included in both samples.

appendix. Since 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.

Percentages of SD and/or LEP students for the national sample of public and nonpublic schools in which accommodations were not permitted are presented in table A.7. The data in this table include the percentages of students *identified* as SD and/or LEP, the percentage of students *excluded*, and the percentage of *assessed* SD and/or LEP students. Tables A.8 and A.9 show similar information by jurisdiction for grade 4 and grade 8. Percentages of these students in the national sample where accommodations were permitted are

presented in table A.10. The state and jurisdiction results where accommodations were permitted are shown in tables A.11 and A.12 for grade 4 and grade 8. The data in these tables include the percentages of students *identified* as SD and/or LEP, the percentage of students *excluded*, the percentage of *assessed* SD and/or LEP students, the percentage *assessed without accommodations*, and the percentage *assessed with accommodations*.

In the 2002 national sample, 6 percent of students at grades 4, 5 percent of students at grade 8, and 4 percent of students at grade 12 were excluded from the assessment (see table A.10). Across the various jurisdictions that participated in the 2002 state assessment, the percentage of students excluded ranged from 3 to 12 percent at grade 4 (see table A.11) and from 2 to 10 percent at grade 8 (see table A.12).

Table A.7 Students with disabilities and/or limited English proficient students identified, excluded, and assessed, when accommodations were not permitted, grades 4, 8, and 12 public and nonpublic schools: 1992–2000

	1992		1994		1998		2000	
	Number of students	Weighted percentage of students sampled	Number of students	Weighted percentage of students sampled	Number of students	Weighted percentage of students sampled	Number of students	Weighted percentage of students sampled
Grade 4								
SD ¹ and/or LEP ² students								
Identified	2,013	10	1,624	13	985	16	823	15
Excluded	1,750	6	1,025	5	545	9	393	7
Assessed	263	4	599	8	440	7	430	8
SD ¹ students								
Identified	1,149	7	1,039	10	490	11	524	11
Excluded	990	4	685	4	247	6	295	6
Assessed	159	3	354	6	243	5	229	5
LEP ² students								
Identified	945	3	623	4	527	6	356	5
Excluded	835	2	368	1	323	3	141	2
Assessed	110	1	255	2	204	2	215	3
Grade 8								
SD ¹ and/or LEP ² students								
Identified	2,310	13	1,737	15	1,365	12	—	—
Excluded	2,030	9	1,278	9	623	6	—	—
Assessed	280	4	459	6	742	7	—	—
SD ¹ students								
Identified	1,522	10	1,323	12	975	10	—	—
Excluded	1,323	7	979	8	524	5	—	—
Assessed	199	3	344	5	451	5	—	—
LEP ² students								
Identified	836	3	444	3	449	3	—	—
Excluded	750	2	323	2	134	1	—	—
Assessed	86	1	121	1	315	2	—	—
Grade 12								
SD ¹ and/or LEP ² students								
Identified	1,547	9	1,237	11	1,011	7	—	—
Excluded	1,417	7	948	7	448	3	—	—
Assessed	130	2	289	4	563	4	—	—
SD ¹ students								
Identified	1,164	7	957	9	669	6	—	—
Excluded	1,088	6	776	6	365	3	—	—
Assessed	76	1	181	3	304	3	—	—
LEP ² students								
Identified	408	2	294	2	392	2	—	—
Excluded	351	1	184	1	115	#	—	—
Assessed	57	1	110	1	277	2	—	—

— Data were not collected at grades 8 and 12 in 2000.

Percentage rounds to zero.

¹ Students with disabilities.

² Limited English proficient students.

NOTE: Within each grade level, the combined SD/LEP portion of the table is not a sum of the separate SD and LEP portions because some students were identified as both SD and LEP. Such students would be counted separately in the bottom portions but counted only once in the top portion. Within each portion of the table, percentages may not add to totals, due to rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, and 2000 Reading Assessments.

Table A.8 Percentage of students with disabilities and/or limited English proficient students identified, excluded, and assessed, when accommodations were not permitted, grade 4 public schools: By state, 1992–1998

Grade 4	1992			1994			1998		
	Identified	Excluded	Assessed	Identified	Excluded	Assessed	Identified	Excluded	Assessed
Nation (Public)	11	6	4	14	6	8	17	10	7
Alabama	10	6	4	11	5	5	13	8	5
Arizona	16	7	9	21	7	14	22	10	12
Arkansas	11	5	6	12	6	6	11	5	6
California	28	14	13	31	12	18	31	15	15
Colorado	11	6	4	15	7	8	15	7	8
Connecticut	15	7	8	17	8	8	18	13	6
Delaware	12	6	6	15	6	9	16	7	9
Florida	17	9	8	22	10	11	18	9	9
Georgia	9	5	4	11	5	5	11	7	4
Hawaii	13	6	8	12	5	7	15	5	10
Idaho	9	4	5	12	5	7	—	—	—
Illinois	—	—	—	—	—	—	14	10	5
Indiana	8	4	3	11	5	6	—	—	—
Iowa	9	4	6	11	5	6	15	8	7
Kansas	—	—	—	—	—	—	12	6	7
Kentucky	8	4	4	8	4	4	13	9	4
Louisiana	8	4	4	11	6	5	15	12	3
Maine	12	5	6	17	10	7	15	8	7
Maryland	14	7	7	15	7	8	13	10	3
Massachusetts	17	7	10	18	8	10	19	8	11
Michigan	7	5	2	10	6	4	10	7	3
Minnesota	10	4	6	12	4	8	15	4	11
Mississippi	7	5	2	9	6	4	7	4	3
Missouri	11	5	6	12	5	7	14	7	7
Montana	—	—	—	11	4	8	10	4	6
Nebraska	13	4	9	16	4	12	—	—	—
Nevada	—	—	—	—	—	—	20	12	7
New Hampshire	12	4	7	15	6	9	14	5	9
New Jersey	10	6	5	12	6	6	—	—	—
New Mexico	13	8	6	18	8	10	28	11	16
New York	13	6	7	15	8	7	14	9	5
North Carolina	12	4	7	14	5	9	15	10	5
North Dakota	10	2	8	10	2	8	—	—	—
Ohio	10	6	4	—	—	—	—	—	—
Oklahoma	13	8	4	—	—	—	15	9	6
Oregon	—	—	—	—	—	—	20	7	12
Pennsylvania	9	4	5	11	6	5	—	—	—
Rhode Island	16	7	9	15	5	10	20	7	12
South Carolina	11	6	5	13	7	6	16	11	5
Tennessee	11	5	7	13	6	6	13	4	9
Texas	17	8	9	24	11	13	26	14	13
Utah	10	4	6	12	5	7	14	5	9
Virginia	12	6	6	13	7	6	15	8	7
Washington	—	—	—	15	5	9	15	5	10
West Virginia	8	5	3	12	7	5	12	9	3
Wisconsin	11	7	4	13	7	6	16	10	6
Wyoming	11	4	7	11	4	7	14	4	9
Other Jurisdictions									
District of Columbia	12	10	3	12	9	3	16	11	6
DDESS ¹	—	—	—	—	—	—	8	5	4
DoDDS ²	—	—	—	9	5	5	7	4	3
Guam	12	7	5	12	9	3	—	—	—
Virgin Islands	6	3	3	—	—	—	8	6	2

— Indicates that the jurisdiction did not participate.

¹ Department of Defense Domestic Dependent Elementary and Secondary Schools.

² Department of Defense Dependents Schools (Overseas).

NOTE: Percentages may not add to totals, due to rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, and 1998 Reading Assessments.

Table A.9 Percentage of students with disabilities and/or limited English proficient students identified, excluded, and assessed, when accommodations were not permitted, grade 8 public schools: By state, 1998

Grade 8	1998		
	Identified	Excluded	Assessed
Nation (Public)	14	6	7
Alabama	12	6	6
Arizona	17	7	11
Arkansas	12	7	5
California	23	8	15
Colorado	14	5	9
Connecticut	15	8	7
Delaware	14	6	8
Florida	17	5	12
Georgia	12	5	7
Hawaii	15	6	9
Illinois	12	6	6
Kansas	12	5	7
Kentucky	10	5	5
Louisiana	14	10	4
Maine	14	7	7
Maryland	12	7	5
Massachusetts	17	7	10
Minnesota	13	4	9
Mississippi	11	7	3
Missouri	13	6	6
Montana	11	3	8
Nevada	15	8	8
New Mexico	22	7	15
New York	16	10	6
North Carolina	14	9	5
Oklahoma	13	9	5
Oregon	14	4	11
Rhode Island	16	5	12
South Carolina	12	6	5
Tennessee	14	4	9
Texas	19	7	12
Utah	11	5	7
Virginia	13	7	6
Washington	13	4	8
West Virginia	14	8	6
Wisconsin	14	8	6
Wyoming	10	2	8
Other Jurisdictions			
District of Columbia	14	9	5
DDESS ¹	10	5	5
DoDDS ²	8	4	4
Virgin Islands	7	7	0

¹ Department of Defense Domestic Dependent Elementary and Secondary Schools.

² Department of Defense Dependents Schools (Overseas).

NOTE: Percentages may not add to totals, due to rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 Reading Assessment.

Table A.10 Students with disabilities and/or limited English proficient students identified, excluded, and assessed, when accommodations were permitted, grades 4, 8, and 12 public and nonpublic schools: 1998–2002

	1998		2000		2002	
	Number of students	Weighted percentage of students sampled	Number of students	Weighted percentage of students sampled	Number of students	Weighted percentage of students sampled
Grade 4						
SD¹ and/or LEP² students						
Identified	973	16	906	18	28,073	19
Excluded	393	6	316	6	10,307	6
Assessed	580	10	590	12	17,766	13
Without accommodations	413	7	476	10	11,913	9
With accommodations	167	3	114	2	5,853	4
SD¹ students						
Identified	558	10	510	11	19,936	12
Excluded	246	4	193	4	8,042	5
Assessed	312	6	317	7	11,894	7
Without accommodations	179	3	209	5	6,631	4
With accommodations	133	3	108	2	5,263	3
LEP² students						
Identified	446	6	446	8	10,334	8
Excluded	167	2	159	3	3,410	2
Assessed	279	4	287	5	6,924	6
Without accommodations	238	3	273	5	6,020	6
With accommodations	41	1	14	#	904	1
Grade 8						
SD¹ and/or LEP² students						
Identified	1,252	12	—	—	20,137	17
Excluded	368	4	—	—	7,135	5
Assessed	884	9	—	—	13,002	11
Without accommodations	678	6	—	—	8,598	8
With accommodations	206	2	—	—	4,404	4
SD¹ students						
Identified	865	10	—	—	16,159	12
Excluded	283	3	—	—	5,939	4
Assessed	582	7	—	—	10,220	8
Without accommodations	404	5	—	—	6,074	5
With accommodations	178	2	—	—	4,146	3
LEP² students						
Identified	447	3	—	—	5,516	6
Excluded	109	1	—	—	1,907	2
Assessed	338	2	—	—	3,609	4
Without accommodations	307	2	—	—	3,113	4
With accommodations	31	#	—	—	496	#

See footnotes at end of table. ►

Table A.10 Students with disabilities and/or limited English proficient students identified, excluded, and assessed, when accommodations were permitted, grades 4, 8, and 12 public and nonpublic schools: 1998–2002 — Continued

	1998		2000		2002	
	Number of students	Weighted percentage of students sampled	Number of students	Weighted percentage of students sampled	Number of students	Weighted percentage of students sampled
Grade 12						
SD¹ and/or LEP² students						
Identified	975	7	—	—	1,556	12
Excluded	327	2	—	—	616	4
Assessed	648	5	—	—	940	8
Without accommodations	532	4	—	—	673	6
With accommodations	116	1	—	—	267	2
SD¹ students						
Identified	649	6	—	—	1,231	9
Excluded	285	2	—	—	535	3
Assessed	364	4	—	—	696	6
Without accommodations	266	3	—	—	446	4
With accommodations	98	1	—	—	250	2
LEP² students						
Identified	353	2	—	—	419	3
Excluded	58	#	—	—	125	1
Assessed	295	2	—	—	294	3
Without accommodations	277	2	—	—	266	2
With accommodations	18	#	—	—	28	#

— Data were not collected at grades 8 and 12 in 2000.

Percentage rounds to zero.

¹ Students with disabilities.

² Limited English proficient students.

NOTE: Within each grade level, the combined SD/LEP portion of the table is not a sum of the separate SD and LEP portions because some students were identified as both SD and LEP. Such students would be counted separately in the bottom portions but counted only once in the top portion.

Within each portion of the table, percentages may not add to totals, due to rounding.

The number of students at grades 4 and 8 are larger in 2002 than in previous years because the 2002 national sample was based on the combined sample of students in each participating state, plus an additional sample from non-participating states as well as a sample from private schools.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998, 2000, and 2002 Reading Assessments.

Table A.11 Percentage of students with disabilities and/or limited English proficient students identified, excluded, and assessed, when accommodations were permitted, grade 4 public schools: By state, 1998 and 2002

Grade 4	1998						2002					
	SD ¹ and/or LEP ² students					All students assessed without accommodations	SD ¹ and/or LEP ² students					All students assessed without accommodations
	Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations		Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations	
Nation (Public)	18	7	11	7	3	90	21	7	14	10	4	89
Alabama	13	8	4	3	1	90	14	3	12	9	2	95
Arizona	22	10	12	10	1	88	28	8	21	18	3	90
Arkansas	11	5	6	4	2	93	14	5	10	8	2	93
California †	31	14	16	15	1	84	34	5	29	28	1	94
Connecticut	18	10	8	5	3	87	16	5	11	5	6	89
Delaware	16	1	15	11	4	95	17	8	9	4	5	87
Florida	18	6	12	8	5	89	25	7	18	10	8	85
Georgia	11	5	6	3	3	93	13	4	9	6	3	93
Hawaii	15	5	10	9	1	94	18	6	12	7	5	89
Idaho	—	—	—	—	—	—	17	4	13	11	2	93
Illinois †	14	6	8	6	2	92	20	7	14	8	6	87
Indiana	—	—	—	—	—	—	13	5	9	7	2	93
Iowa †	15	5	10	7	3	92	16	8	8	3	5	87
Kansas †	12	4	8	5	4	93	19	5	14	7	7	88
Kentucky	13	7	5	3	2	90	12	8	4	3	1	91
Louisiana	15	7	8	3	5	88	19	10	9	3	6	84
Maine	15	7	7	4	3	90	17	6	11	5	6	88
Maryland	13	6	8	4	4	90	14	7	7	5	2	92
Massachusetts	19	5	14	9	5	90	19	6	13	4	9	85
Michigan	10	6	4	3	1	93	14	7	6	5	1	92
Minnesota †	15	3	12	9	3	94	19	5	13	10	4	91
Mississippi	7	4	3	2	#	95	7	4	3	2	1	95
Missouri	14	6	8	3	4	89	16	9	8	4	3	88
Montana †	10	2	7	5	2	96	15	6	8	4	4	89
Nebraska	—	—	—	—	—	—	21	5	15	9	6	88
Nevada	20	11	9	8	1	88	27	10	17	14	3	87
New Mexico	28	9	18	16	2	88	37	10	27	23	4	85
New York †	14	7	7	2	4	88	18	8	9	3	6	86
North Carolina	15	7	9	3	6	88	19	12	7	3	4	84
North Dakota †	—	—	—	—	—	—	18	5	13	9	3	91
Ohio	—	—	—	—	—	—	14	8	5	4	2	90
Oklahoma	15	9	6	5	1	90	21	5	15	10	5	89
Oregon	20	6	14	10	4	90	25	8	17	13	4	88
Pennsylvania	—	—	—	—	—	—	14	5	10	4	5	90
Rhode Island	20	7	13	9	4	89	25	6	19	8	11	84
South Carolina	16	8	9	6	3	90	16	5	12	9	3	92
Tennessee †	13	4	9	8	2	95	14	3	10	9	1	95
Texas	26	13	14	11	3	85	27	11	16	14	2	87
Utah	14	6	8	6	2	92	19	6	13	9	4	91
Vermont	—	—	—	—	—	—	15	5	10	4	6	89
Virginia	15	6	9	4	5	89	18	10	8	5	3	87
Washington †	15	5	10	7	3	92	15	5	11	7	4	92
West Virginia	12	8	4	2	1	90	16	10	5	3	2	87
Wisconsin †	16	8	8	5	3	89	19	8	10	5	5	87
Wyoming	14	3	10	6	4	93	17	3	15	7	7	90
Other Jurisdictions												
District of Columbia	16	9	8	5	3	89	19	8	11	5	5	86
DDESS ³	8	4	4	2	2	94	14	4	10	6	4	92
DoDDS ⁴	7	3	4	3	1	96	16	3	13	9	4	93
Guam	—	—	—	—	—	—	39	7	32	26	6	87
Virgin Islands	8	5	3	2	1	94	7	3	4	4	1	97

— Indicates that the jurisdiction did not participate.

Percentage rounds to zero.

† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2002.

¹ Students with disabilities. ² Limited English proficient students.

³ Department of Defense Domestic Dependent Elementary and Secondary Schools. ⁴ Department of Defense Dependents Schools (Overseas).

NOTE: Percentages may not add to totals, due to rounding.

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

Table A.12 Percentage of students with disabilities and/or limited English proficient students identified, excluded, and assessed, when accommodations were permitted, grade 8 public schools: By state, 1998 and 2002

Grade 8	1998						2002					
	SD ¹ and/or LEP ² students					All students assessed without accommodations	SD ¹ and/or LEP ² students					All students assessed without accommodations
	Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations		Identified	Excluded	Assessed	Assessed without accommodations	Assessed with accommodations	
Nation (Public)	14	4	10	7	3	93	18	6	12	8	4	90
Alabama	12	6	6	5	#	93	14	2	12	11	1	97
Arizona	17	5	12	10	1	93	21	5	16	14	2	93
Arkansas	12	5	6	5	1	94	15	5	10	9	2	93
California †	23	4	19	17	2	94	26	4	23	21	2	94
Connecticut	15	6	9	7	3	91	17	4	12	6	6	90
Delaware	14	2	13	10	2	96	15	6	9	2	6	88
Florida	17	5	12	9	3	92	21	6	15	8	8	86
Georgia	12	4	8	5	3	93	13	4	8	5	3	93
Hawaii	15	5	10	7	3	92	20	5	15	10	5	90
Idaho	—	—	—	—	—	—	14	4	10	8	2	94
Illinois †	12	4	8	6	3	93	16	4	13	7	6	90
Indiana	—	—	—	—	—	—	14	4	11	7	3	93
Kansas †	12	4	8	6	2	95	16	5	11	6	5	90
Kentucky	10	3	6	4	3	94	12	7	5	4	1	92
Louisiana	14	5	9	4	5	90	16	10	6	3	3	87
Maine	14	5	9	6	3	92	17	4	13	8	6	90
Maryland	12	3	9	3	5	92	15	4	10	8	2	93
Massachusetts	17	4	12	8	5	91	20	6	14	6	8	86
Michigan	—	—	—	—	—	—	13	7	6	4	2	91
Minnesota †	13	1	12	9	3	96	15	3	12	9	3	94
Mississippi	11	6	5	4	1	94	10	5	5	3	1	93
Missouri	13	4	9	6	3	93	15	8	8	4	4	88
Montana †	11	4	8	6	1	95	13	4	9	7	2	94
Nebraska	—	—	—	—	—	—	17	7	10	7	2	91
Nevada	15	6	9	8	2	92	20	6	14	12	2	92
New Mexico	22	8	14	10	4	88	31	8	23	17	5	86
New York †	16	8	8	3	5	88	20	9	11	4	7	83
North Carolina	14	6	8	3	5	89	18	9	9	3	6	85
North Dakota †	—	—	—	—	—	—	15	4	11	8	2	93
Ohio	—	—	—	—	—	—	12	7	5	4	1	91
Oklahoma	13	9	4	4	1	90	17	4	13	10	4	92
Oregon †	14	4	10	6	4	92	18	5	13	10	3	92
Pennsylvania	—	—	—	—	—	—	15	3	12	4	8	89
Rhode Island	16	6	10	9	1	92	20	5	15	8	7	88
South Carolina	12	5	7	5	1	93	14	5	9	6	3	92
Tennessee †	14	6	8	7	1	93	13	3	9	9	1	96
Texas	19	5	13	11	3	92	20	8	12	11	1	91
Utah	11	4	7	6	2	95	15	4	11	9	2	94
Vermont	—	—	—	—	—	—	18	5	13	8	6	89
Virginia	13	5	8	4	3	91	17	8	9	5	4	88
Washington †	13	4	9	6	3	94	14	4	10	6	5	92
West Virginia	14	7	7	4	2	90	16	10	7	4	2	88
Wisconsin †	14	5	9	5	4	91	16	7	9	4	5	88
Wyoming	10	2	8	7	1	96	14	3	11	6	6	91
Other Jurisdictions												
American Samoa	—	—	—	—	—	—	22	8	14	10	4	88
District of Columbia	14	5	9	6	3	92	21	7	13	5	8	84
DDESS ³	10	2	9	5	4	95	13	3	10	5	5	92
DoDDS ⁴	8	1	7	5	2	97	10	2	8	6	3	96
Guam	—	—	—	—	—	—	29	2	27	25	3	95
Virgin Islands	7	7	0	0	0	93	11	8	3	3	#	91

— Indicates that the jurisdiction did not participate.

Percentage rounds to zero.

† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2002.

¹ Students with disabilities. ² Limited English proficient students.

³ Department of Defense Domestic Dependent Elementary and Secondary Schools. ⁴ Department of Defense Dependents Schools (Overseas).

NOTE: Percentages may not add to totals, due to rounding.

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

Investigating the Potential Effects of Exclusion Rates on Assessment Results

Since students with disabilities or limited English proficient students tend to score below average on assessments, excluding students with special needs may increase a jurisdiction's scores. Conversely, including more of these students might depress score gains. In 2002, exclusion rates varied among jurisdictions. In addition, cases of both increases and decreases in exclusion rates occurred between 1998 and 2002, making comparisons over time within jurisdictions complex to interpret. Thus, the potential impact of exclusion rates on assessment results is a validity concern. 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. Tables A.11 and A.12 on the preceding pages display the rates of exclusion in 1998 and 2002 in each jurisdiction for grade 4 and grade 8, respectively.

As shown in table A.13, of the 48 jurisdictions that assessed reading at grade 4 in 2002, seven jurisdictions had exclusion rates of 10 percent or greater, while the majority had exclusion rates of less than eight percent. Table A.14 displays the comparable data for grade 8. Seven jurisdictions at grade 8 had exclusion rates of 8 percent or above, although none was above 10 percent. The other jurisdictions at grade 8 all had exclusion rates of less than 8 percent.

Table A.13 Grouping of states/jurisdictions by percentage of excluded students in 2002: Grade 4

Grade 4	Percentage excluded	Number of states/ jurisdictions	States/jurisdictions	
	0–4.9%	16	Alabama Arkansas Connecticut DDESS ¹ DoDDS ² Georgia Idaho Indiana	Mississippi Pennsylvania South Carolina Tennessee Vermont Virgin Islands Washington Wyoming
	5–7.9%	19	Arizona California Florida Guam Hawaii Iowa Kansas Maine Maryland Massachusetts	Michigan Minnesota Montana Nebraska North Dakota Oklahoma Oregon Rhode Island Utah
	8–9.9%	6	Delaware District of Columbia Kentucky	Missouri New York Ohio
	10% or Greater	7	Louisiana Nevada New Mexico North Carolina	Texas Virginia West Virginia

¹ Department of Defense Domestic Dependent Elementary and Secondary Schools.

² Department of Defense Dependents Schools (Overseas).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Table A.14 Grouping of states/jurisdictions by percentage of excluded students in 2002: Grade 8

Grade 8	Percentage excluded	Number of states/ jurisdictions	States/jurisdictions	
	0–4.9%	22	Alabama Arkansas California Connecticut DDESS ¹ DoDDS ² Georgia Guam Hawaii Idaho Indiana	Maine Maryland Montana North Dakota Oklahoma Pennsylvania Tennessee Utah Vermont Washington Wyoming
	5–7.9%	18	American Samoa Arizona Delaware District of Columbia Florida Kansas Kentucky Massachusetts Michigan	Mississippi Missouri Nebraska Nevada Ohio Oregon Rhode Island South Carolina Virginia
	8–9.9%	7	Louisiana New Mexico New York North Carolina	Texas Virgin Islands West Virginia

¹ Department of Defense Domestic Dependent Elementary and Secondary Schools.

² Department of Defense Dependents Schools (Overseas).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

There is variability in exclusion rates across states due to at least three factors. One factor is that the percentage of students who are *identified* as having disabilities or limited proficiency in English varies across jurisdictions and over time. Reasons for this variation include: 1) lack of standardized criteria for defining students as having specific disabilities or as being limited in their English proficiency; 2) changes or differences in policy and practices regarding implementation of the Individuals with Disabilities Education Act

(IDEA); and 3) population shifts in the percentage of students classified as limited English proficient and, to a lesser extent, as students with disabilities.

The second factor is that some SD and/or LEP students are excluded because they require accommodations, such as testing in another language or reading the passage aloud, that would be inconsistent with NAEP’s reading framework and would change the construct that NAEP intends to measure.

The third factor is that some SD and/or LEP students are excluded because they are so severely disabled or lacking in English language skills that no accommodation would be sufficient to enable them to meaningfully participate.

With regard to cross-state comparisons, the correlations between rates of exclusion and average 2002 reading scores were not found to be significant at either grade 4 (.05) or grade 8 (-.21). In other words, higher exclusion rates were not associated with higher average scores in 2002. However, with regard to state trends, the correlations between changes in the rate of exclusion of students with special needs and average reading scores gains from 1998 to 2002 were found to be moderate (.50 at grade 4 and .56 at grade 8). While there was a moderate tendency for an increase in exclusion rates to be associated with an increase in average scale scores, exclusion increases do not explain the entirety of score gains.

Because the representativeness of samples is ultimately a validity issue, NCES has commissioned studies of the impact of assessment accommodations on overall scores. NCES has also investigated scenarios for estimating what the average scores might have been had the excluded students been assessed. Several statistical scenarios have been proposed, based on different hypotheses about how excluded students might have performed. Combined with the actual performance of students who were assessed, these scenarios produce results for the full population (that is, including estimates for excluded students) in each jurisdiction and each assessment year. Although these scenarios are somewhat speculative, these techniques do provide some indication as to which statements about trend gains or losses *might* be changed if exclusion rates were zero in both assessment years and if the hypotheses about the performance of missing students are correct.

Although the results of one of these scenarios are presented below, the methods used to construct the scenario are still under development. NCES is continuing research into different procedures for reducing the percentages of students excluded from NAEP. In addition, NCES will continue to evaluate the potential impact of changes in exclusion rates on score gains. The scenario shown in this appendix is provided to illustrate the potential impact of reasonable hypotheses about the performance of excluded students on score gains in the jurisdictions that participated in both 1998 and 2002 and should not be interpreted as official results.

The scenario was developed by Donald McLaughlin of American Institutes for Research, and predicts what the performance of excluded SD and/or LEP students might have been had these students been tested. The basic assumption underlying this approach is that these students would have performed as well as included SD and/or LEP students with similar disabilities, level of English proficiency, and background characteristics.⁷ The scenario was performed for each jurisdiction that participated in both 1998 and 2002.

The first column of table A.15 presents the official grade 4 score gain (or loss) for each jurisdiction based on the results shown in table 2.2 in chapter 2 of this report. The second column shows the score gain (or loss) under the McLaughlin scenario. Five jurisdictions have notations that show that a trend reported as significant or as not significant would change under this scenario. For example, in Arkansas the apparent score gain between 1998 and 2002 of 4.1 points was not statistically significant, but under this scenario, the hypothetical gain of 5.3 points would have been significant. The third column reports the difference between the official gain and the gain under this scenario. For Arkansas, this difference is 1.2 points. Similar data are presented for grade 8 in table A.16. At grade 8, five states and jurisdictions have notations indicating that the trend reported as significant or as not significant would change under this scenario.

⁷ Because students with very severe levels of disability and students with little or no proficiency in English are not assessed in NAEP, ability estimates for students with those characteristics may be overestimated.

Table A.15 Comparison of changes in average NAEP reading scores from 1998 to 2002 in the official NAEP reported sample and one possible scenario that includes estimates of how excluded students might have performed had they been assessed: Grade 4

Grade 4	Reported sample	Scenario ¹	Difference in score change (Scenario minus reported)
Alabama	-4.5	-1.3	3.2
Arizona	-1.1	1.9	3.0
Arkansas ²	4.1	5.3	1.2
California ^{2†}	3.5	8.9	5.4
Connecticut	-0.6	2.7	3.3
Delaware	17.3	15.0	-2.3
District of Columbia	11.3	12.4	1.1
Florida	8.7	8.6	-0.1
Georgia	6.3	7.0	0.7
Hawaii	7.9	7.4	-0.5
Iowa [†]	3.2	2.6	-0.6
Kansas [†]	0.7	0.4	-0.4
Kentucky	1.5	0.2	-1.3
Louisiana	6.3	7.4	1.1
Maine	-0.3	0.4	0.7
Maryland	5.5	5.6	0.1
Massachusetts	10.9	12.1	1.1
Michigan	2.7	2.6	0.0
Minnesota [†]	6.0	6.1	0.1
Mississippi	-0.4	0.5	0.9
Missouri	4.6	3.8	-0.8
Montana [†]	-0.6	-1.7	-1.0
Nevada ²	3.3	5.6	2.3
New Mexico	2.6	4.2	1.6
New York [†]	7.0	7.4	0.4
North Carolina	8.7	9.7	1.0
Oklahoma ²	-5.9	-3.3	2.5
Oregon	8.4	8.7	0.4
Rhode Island	1.7	3.1	1.3
South Carolina	5.0	6.8	1.8
Tennessee [†]	1.9	2.4	0.5
Texas	2.8	4.6	1.8
Utah	5.3	6.0	0.7
Virgin Islands ²	5.4	7.9	2.5
Virginia	7.8	7.4	-0.4
Washington [†]	5.5	5.4	-0.1
West Virginia	3.2	3.2	0.1
Wyoming	2.9	3.3	0.4

[†] Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2002.

¹ This scenario assumes that excluded SD and/or LEP students would have performed as well as assessed SD and/or LEP students with similar special needs.

² The official reported 1998 vs. 2002 trend results for this state would be different under the scenario.

NOTE: Only states or jurisdictions that participated in both 1998 and 2002 reading assessments are presented in this table. Scenario results are not available for the Department of Defense Schools.

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

Table A.16 Comparison of changes in average NAEP reading scores from 1998 to 2002 in the official NAEP reported sample and one possible scenario that includes estimates of how excluded students might have performed had they been assessed: Grade 8

Grade 8	Reported sample	Scenario ¹	Difference in score change (Scenario minus reported)
Alabama	-2.5	-0.5	2.0
Arizona	-3.2	-3.1	0.1
Arkansas	4.1	4.3	0.2
California †	-1.9	-1.2	0.7
Connecticut ²	-3.4	-2.2	1.2
Delaware	13.6	11.2	-2.3
District of Columbia	4.1	2.9	-1.2
Florida	6.5	6.8	0.3
Georgia	0.8	1.0	0.3
Hawaii	2.8	3.0	0.2
Kansas †	1.5	0.7	-0.7
Kentucky	2.9	1.8	-1.1
Louisiana ²	4.8	3.5	-1.3
Maine	-1.6	-0.8	0.8
Maryland	2.4	2.1	-0.3
Massachusetts	1.7	1.6	-0.2
Mississippi	3.6	4.8	1.3
Missouri	5.6	4.1	-1.4
Montana †	-0.8	-0.9	-0.1
Nevada	-6.4	-5.9	0.5
New Mexico	-4.3	-4.8	-0.5
New York †	-0.8	-0.6	0.2
North Carolina	2.7	2.9	0.2
Oklahoma ²	-3.2	0.6	3.9
Oregon †	2.1	1.7	-0.5
Rhode Island	-2.5	-1.3	1.2
South Carolina	2.8	3.2	0.3
Tennessee ^{2†}	2.3	4.9	2.6
Texas	0.9	-0.3	-1.3
Utah	-0.1	0.1	0.3
Virgin Islands ²	9.6	4.8	-4.8
Virginia	2.7	2.0	-0.7
Washington †	4.4	5.2	0.8
West Virginia	1.9	0.9	-1.0
Wyoming	1.7	1.7	0.0

† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2002.

¹ This scenario assumes that excluded SD and/or LEP students would have performed as well as assessed SD and/or LEP students with similar special needs.

² The official reported 1998 vs. 2002 trend results for this state would be different under the scenario.

NOTE: Only states or jurisdictions that participated in both 1998 and 2002 reading assessments are presented in this table. Scenario results are not available for the Department of Defense Schools.

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

Table A.17 displays jurisdictions by the size of the difference between the reported grade 4 gains in average scores and the gains under this scenario. For 20 of the 38 jurisdictions that participated in both 1998 and 2002 fourth-grade reading assessments (and for which the scenario results are available), the scenario would make no more than one scale point difference one way or the other. Of the 38 jurisdictions, 35 might have differed by less than three points. Three jurisdictions might have differed by three points or more.

Table A.18 displays the same information for grade 8. For 22 of the 35 jurisdictions that participated in both 1998 and 2002 eighth-grade reading assessments (and for which the scenario results are available), the scenario would make no more than one scale point difference one way or the other. Thirty-three of the 35 jurisdictions might have differed by up to three points, and two additional jurisdictions might have differed by more than three points.

Table A.17 Frequency distribution of differences between Reported and Scenario¹ average score changes from 1998 to 2002: Grade 4

Grade 4	Difference in score change (Scenario minus reported)	Number of states/jurisdictions	States/jurisdictions
	–3.00 to –1.01	3	Delaware, Kentucky, Montana
	–1.00 to 0.99	20	Florida, Georgia, Hawaii, Iowa, Kansas, Maine, Maryland, Michigan, Minnesota, Mississippi, Missouri, New York, North Carolina, Oregon, Tennessee, Utah, Virginia, Washington, West Virginia, Wyoming
	1.00 to 2.99	12	Arizona, Arkansas, ² District of Columbia, Louisiana, Massachusetts, Nevada, ² New Mexico, Oklahoma, ² Rhode Island, South Carolina, Texas, Virgin Islands ²
	3.00 to 4.99	2	Alabama, Connecticut
	5.00 to 5.99	1	California ²

¹ The scenario assumes that all excluded SD and/or LEP students would have performed as well as assessed SD and/or LEP students with similar special needs.

² The official reported 1998 vs. 2002 trend results for this state would be different under the scenario.

NOTE: Only states or jurisdictions that participated in both 1998 and 2002 reading assessments are presented in this table. Scenario results are not available for the Department of Defense Schools.

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

Table A.18 Frequency distribution of differences between Reported and Scenario¹ average score changes from 1998 to 2002: Grade 8

Grade 8	Difference in score change (Scenario minus reported)	Number of states/jurisdictions	States/jurisdictions
	–6.00 to –3.01	1	Virgin Islands ²
	–3.00 to –1.01	6	Delaware, District of Columbia, Kentucky, Louisiana ² , Missouri, Texas
	–1.00 to 0.99	22	Arizona, Arkansas, California, Florida, Georgia, Hawaii, Kansas, Maine, Maryland, Massachusetts, Montana, Nevada, New Mexico, New York, North Carolina, Oregon, South Carolina, Utah, Virginia, Washington, West Virginia, Wyoming
	1.00 to 2.99	5	Alabama, Connecticut ² , Mississippi, Rhode Island, Tennessee ²
	3.00 to 4.99	1	Oklahoma ²

¹ The scenario assumes that all excluded SD and/or LEP students would have performed as well as assessed SD and/or LEP students with similar special needs.

² The official reported 1998 vs. 2002 trend results for this state would be different under the scenario.

NOTE: Only states or jurisdictions that participated in both 1998 and 2002 reading assessments are presented in this table. Scenario results are not available for the Department of Defense Schools.

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

Types of Accommodations Permitted

Table A.19 displays the percentages of SD and/or LEP students assessed with the variety of 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. 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, however, 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.

Table A.19 Students with disabilities and/or limited English proficient students assessed with accommodations, by type of primary accommodation, grades 4, 8, and 12 public and nonpublic schools: 1998–2002

	Weighted percentage of students sampled						
	Grade 4			Grade 8		Grade 12	
	1998	2000	2002	1998	2002	1998	2002
SD¹ and/or LEP² students							
Large-print book	0.00	0.06	0.04	0.14	0.01	0.04	0.01
Extended time	1.07	0.86	1.65	1.07	2.08	0.39	1.27
Small group	1.94	1.48	2.18	1.26	1.64	0.66	0.73
One-on-one	0.23	0.27	0.09	0.07	0.05	0.15	0.03
Scribe/computer	0.05	0.03	0.06	0.00	0.03	0.00	0.00
Other	0.09	0.01	0.04	0.00	0.04	0.05	0.07
SD students only							
Large-print book	0.00	0.06	0.04	0.14	0.01	0.04	0.01
Extended time	0.78	0.86	1.32	0.86	1.85	0.34	1.18
Small group	1.70	1.36	2.04	1.25	1.57	0.60	0.73
One-on-one	0.23	0.27	0.08	0.07	0.05	0.14	0.03
Scribe/computer	0.05	0.03	0.06	0.00	0.03	0.00	0.00
Other	0.09	0.01	0.03	0.00	0.04	0.02	0.07
LEP² students only							
Large-print book	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Extended time	0.31	0.01	0.44	0.23	0.38	0.05	0.17
Small group	0.32	0.20	0.25	0.01	0.14	0.07	0.01
One-on-one	0.00	0.01	0.01	0.00	0.00	0.01	0.00
Scribe/computer	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Other	0.00	0.01	0.01	0.00	0.00	0.03	0.00

¹ Students with disabilities.

² Limited English proficient students.

NOTE: The combined SD/LEP portion of the table is not a sum of the separate SD and LEP portions because some students were identified as both SD and LEP. Such students would be counted separately in the bottom portions but counted only once in the top portion.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998, 2000, and 2002 Reading Assessments.

Data Collection and Scoring

The 2002 reading assessment was conducted from January to March 2002. Data collection for the 2002 assessment at both the national and state levels was conducted by trained field staff from Westat.

Materials from the 2002 assessment were shipped to NCS Pearson, where trained staff evaluated the responses to the constructed-response questions using scoring rubrics or guides prepared by ETS. Each constructed-response question had a unique scoring guide that defined the criteria used to evaluate students' responses. The extended constructed-response questions were evaluated with four-level guides. The short constructed-response questions were scored as either acceptable or unacceptable or were rated according to three-level guides that permitted partial credit.

For the 2002 reading assessment, 4,023,861 constructed responses were scored. This number includes rescoring to monitor interrater reliability. The within-year average percentage of exact agreement for the 2002 national reliability sample was 92 percent at fourth grade, 91 percent at eighth grade, and 90 percent at twelfth grade.

Data Analysis and IRT Scaling

Subsequent to 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 sample reflected the probability of selection for each student as a result of the sampling design, adjusted for nonresponse.⁸

The procedure used for sample weighting in the state assessments is similar to that used in national samples. However, there is one important difference: because there is no oversampling of high-minority schools in state samples, the weighting process does not need to adjust for such a procedure.

Analyses were then conducted to determine the percentages of students who gave various responses to each cognitive and background question. In determining these percentages for the cognitive questions, a distinction was made between missing responses at the end of a block (i.e., missing responses subsequent to the last question the student answered) and missing responses prior to the last observed response. Missing responses before the last observed response were considered intentional omissions. In analysis, omitted responses to multiple-choice items were scored as fractionally correct.⁹ For constructed-response items, omitted responses were placed into the lowest score category. Missing responses at the end of the block were considered "not reached" and treated as if the questions had not been presented to the student. In calculating response percentages for each question, only students classified as having been presented the question were included in the denominator of the statistic.

⁸ Weighting procedures are described more fully in the "Weighting and Variance Estimation" section later in this document. Additional information about the use of weighting procedures, will be included in the technical documentation section of the NAEP web site at <http://nces.ed.gov/nationsreportcard>.

⁹ Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*, p. 229. Hillsdale, NJ: Lawrence Erlbaum Associates.

It is standard NAEP practice to treat all nonrespondents to the last question in a block as if they had not reached the question. For multiple-choice and short constructed-response questions, this practice produces a reasonable pattern of results in that the proportion reaching the last question is not dramatically smaller than the proportion reaching the next-to-last question. However, for reading blocks that ended with extended constructed-response questions, the standard practice could result in extremely large drops in the proportion of students attempting some of the final questions. Therefore, for blocks ending with an extended constructed-response question, students who answered the next-to-last question but did not respond to the extended constructed-response question were classified as having intentionally omitted the last question.

Item Response Theory (IRT) was used to estimate average reading scale scores for the nation and for various subgroups of interest within the nation. IRT models the probability of answering a question in a certain way as a mathematical function of proficiency or skill. The main purpose of IRT analysis is to provide a common scale on which performance can be compared among groups such as those defined by characteristics, including gender and race/ethnicity, even when students receive different blocks of items. 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 to the assessment items that function at a particular point on the scale

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 1992, 1994, 1998, 2000 and 2002 are presented on the NAEP reading scales. In 1992, a scale ranging from 0 to 500 was created to report performance for each reading purpose — literary and information at grade 4; and literary, information, and task at grades 8 and 12. The scales summarize student performance across all three types of questions in the assessment (multiple-choice, short constructed-response, and extended constructed-response). Results from subsequent reading assessments (1994, 1998, 2000, and 2002) are reported on these scales.

Each reading scale was initially based on the distribution of student performance across all three grades in the 1992 national assessment (grades 4, 8, and 12). In that year, the scales had an average of 250 and a standard deviation of 50. In addition, a composite scale was created as an overall measure of students' reading performance. This composite scale is a weighted average of the three separate scales for the three reading purposes. The weight for each reading purpose is proportional to the relative importance assigned to the reading purpose by the specifications developed through the consensus planning process and given in the framework.

In producing the reading scales, three distinct IRT models were used. Multiple-choice questions were scaled using the three-parameter logistic (3PL) model; short constructed-response questions rated as

¹⁰ Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, 16(2), 159–176.

acceptable or unacceptable were scaled using the two-parameter logistic (2PL) model; and short constructed-response questions rated according to a three-level guide, as well as extended constructed-response questions rated on a four- or five-level guide, were scaled using a Generalized Partial-Credit (GPC) model.¹⁰ Developed by ETS and 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 these more complex constructed-response questions.¹¹

The reading scale is composed of three types of questions: multiple-choice, short constructed-response (scored either dichotomously or allowing for partial credit), and extended constructed-response (scored according to a partial-credit model). Unfortunately, the question of how much information different types of questions contribute to the reading scale has no simple answer. The information provided by a given question is determined by the IRT model used to scale the question. It is a function of the item parameters and varies by level of reading proficiency.¹² Thus, the answer to the query “How much information do the different types of questions provide?” will differ for each level of reading performance. When considering the composite reading scale, the answer is even more complicated. The reading data are scaled separately by the two purposes for reading (reading for information and reading for literary experience) for grade 4,

and the three purposes for reading (reading for information; reading for literary experience; and reading to perform a task) for grades 8 and 12, resulting in two or three separate subscales at each grade. The composite scale is a weighted combination of these subscales. IRT information functions are only strictly comparable when the item parameters are estimated together. Because the composite scale is based on three separate estimation runs, there is no direct way to compare the information provided by the questions on the composite scale.

Because of the PBIB spiraling design used by NAEP, students do not receive enough questions about a specific topic to provide reliable information about individual performance. (For more information on PBIB spiraling, see “The Assessment Design” section presented earlier in this appendix.) Traditional test scores for individual students, even those based on IRT, would result in misleading estimates of population characteristics, such as subgroup means and percentages of students at or above a certain scale-score level. However, it is NAEP’s goal to estimate these population characteristics. As discussed by Mislevy and Sheehan (1987)¹³, 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. Under the assumptions of the scaling models, these population estimates

¹¹ More detailed information regarding the IRT analyses used in NAEP will be included in the technical documentation section of the NAEP web site at <http://nces.ed.gov/nationsreportcard>.

¹² Donoghue, J. R. (1994). An Empirical Examination of the IRT Information of Polytomously Scored Reading Items Under the Generalized Partial Credit Model. *Journal of Educational Measurement*, 31(4), 295–311.

¹³ 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*. Report, No. 15-TR-20, pp. 293-260. Princeton, NJ: Educational Testing Service.

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.¹⁴

Item Mapping Procedures

The reading performance of fourth-, eighth-, and twelfth-graders can be illustrated by “item maps,” which position question or “item” descriptions along the NAEP reading scale at each grade. Each question shown is placed at the point on the scale where questions are likely to be answered successfully by students. The descriptions used on these item maps focus on the reading knowledge or skill needed to answer the question. For multiple-choice questions, the description indicates the knowledge or skill demonstrated by selection of the correct option; for constructed-response questions, the description takes into account the knowledge or skill specified by the different levels of scoring criteria for that question.

To map questions to particular points on the NAEP reading scale, a response probability convention was adopted that would divide those who had a higher probability of success from those who had a lower probability. Establishing a response probability convention has an impact on the mapping of the test questions onto the reading scale. A lower boundary convention maps the reading questions at lower points along the scale, and a higher boundary convention maps the same questions at higher points on the scale. The underlying distribution of reading skills in the population does not change, but the choice of a response probability convention does have

an impact on the proportion of the student population that is reported as “able to do” the questions on the reading scales.

There is no obvious choice of a point along the probability scale that is clearly superior to any other point. If the convention were set with a boundary at 50 percent, those above the boundary would be more likely to get a question right than get it wrong, while those below the boundary would be more likely to get the question wrong than right. Although this convention has some intuitive appeal, it was rejected on the grounds that having a 50:50 chance of getting the question right shows an insufficient degree of mastery. If the convention were set with a boundary at 80 percent, students above the criterion would have a high probability of success with a question. However, many students below this criterion show some level of reading ability that would be ignored by such a stringent criterion. In particular, those in the range between 50 and 80 percent correct would be more likely to get the question right than wrong, yet would not be in the group described as “able to do” the question.

In a compromise between the 50 percent and the 80 percent conventions, NAEP has adopted two related response probability conventions for all its subjects: 65 percent for constructed-response questions (where guessing is not a factor) and 74 percent for multiple-choice questions (to adjust for the possibility of answering correctly by guessing). These probability conventions were established, in part, based on an intuitive judgment that they would provide the best picture of students’ reading skills.

¹⁴ 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.

Some additional support for the dual conventions adopted by NAEP was provided by Huynh.¹⁵ He examined the IRT information provided by items, according to the IRT model used in scaling NAEP questions. Following Bock, Huynh decomposed the item information into that provided by a correct response [$P(\theta) I(\theta)$] and that provided by an incorrect response [$(1 - P(\theta)) I(\theta)$].¹⁶ Huynh showed that the item information provided by a correct response to a constructed-response item is maximized at the point along the reading scale at which the probability of a correct response is 0.65 (for multiple-choice items, the information provided by a correct response is maximized at the point at which the probability of getting the item correct is 0.74). It should be noted, however, that maximizing the item information $I(\theta)$, rather than the information provided by a correct response [$P(\theta) I(\theta)$], would imply an item mapping criterion closer to 50 percent.

The results in this report are presented in terms of the composite reading scale. However, the reading assessment was scaled separately for the two purposes for reading at grade 4 and the three purposes for reading at grades 8 and 12. The composite scale is a weighted combination of the two or three subscales for the two or three purposes for reading. To obtain item map information, a procedure developed by Donoghue was used.¹⁷ This method models the relationship between the item response function for the subscale and the subscale structure to derive the relationship between

the item score and the composite scale (i.e., an item response function for the composite scale). This item response function is then used to derive the probability used in the mapping.

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.

Prior to 2002, the national samples used weights that had been poststratified to the Census or Current Population Survey (CPS) totals for the populations being assessed. There were concerns about the availability of appropriate targets for poststratification

¹⁵ Huynh, H. (1994, October). *Some Technical Aspects of Standard Setting*. Paper presented at the Joint Conference on Standard Setting for Large-Scale Assessment, Washington, DC.

¹⁶ Bock, R. D. (1972). Estimating Item Parameters and Latent Ability When Responses are Scored in Two or More Latent Categories. *Psychometrika*, 37, 29–51.

¹⁷ Donoghue, J. R. (1997, March). *Item Mapping to a Weighted Composite Scale*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

in the 2002 assessment and in the future due to changes in the reporting of race in the 2000 Census. Therefore, in 2002, it was decided that in the analysis of national samples non-poststratified weights would be used. In linking the 2002 NAEP reading results to the existing NAEP reading reporting scale, non-poststratified weights were used throughout the process. This resulted in a slight change to the 1998 National Reading and 2000 National Reading NAEP achievement scores that had been reported previously. The NAEP state samples have always been analyzed using non-poststratified weights since there were no targets available from CPS to use in poststratification. There were no changes to the reported 1998 NAEP state reading achievement results due to this change in the sample weighting procedures.

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: 1) the uncertainty due to sampling only a relatively small number of students, and 2) the uncertainty due to sampling only a portion of the cognitive domain of interest. The first component accounts for the variability associated with the estimated percentages of students who had certain background characteristics or who answered a certain cognitive question correctly.

Because NAEP uses complex sampling procedures, conventional formulas for estimating sampling variability that assume simple random sampling are inappropriate.

NAEP uses a jackknife replication procedure to estimate standard errors. The jackknife standard error provides a reasonable measure of uncertainty for any student information that can be observed without error. However, because each student typically responds to only a few questions within any theme of reading, 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 groups and subgroups 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.¹⁸

Typically, when the standard error is based on a small number of students or when the group of students is enrolled in a small number of schools, the amount of uncertainty associated with the estimation of standard errors may be quite large. Estimates of standard errors subject to a large degree of uncertainty are followed by the “!” symbol to indicate that the nature of the sample does not allow accurate determination of the variability of the statistic. In such cases, the standard errors—and any confidence intervals or significance tests involving these standard errors—should be interpreted cautiously. Additional details concerning procedures for identifying such standard errors will be discussed in the technical documentation section of the NAEP web site at <http://nces.ed.gov/nationsreportcard>.

¹⁸ 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.

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 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 related to the fact that they cannot ask all 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, and observed similarities or differences should not be relied on solely. There-

fore, 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 presented in this report, all the estimates have corresponding estimated standard errors of the estimates. For example, table A.20 shows the average national scale score for the NAEP 1992–2002 national assessments and table A.21 shows 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. Additional examples of estimated standard errors corresponding with results included in this report are presented in tables A.22, A.23, and A.24. For the estimated standard errors corresponding to other data in this report, the reader can go to the data tool on the NCES web site 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 reading 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:

$$\begin{aligned} &\text{Average 1.96 standard errors} \\ &256 \pm 1.96 \times 1.2 \\ &256 \pm 2.4 \\ &(253.6, 258.4) \end{aligned}$$

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 confidence intervals can be constructed for percentages, if the percentages are not extremely large or extremely small. Extreme percentages should be interpreted with caution. Adding or subtracting the standard errors associated with extreme percentages could cause the confidence interval to exceed 100 percent or fall below 0 percent, resulting in numbers that are not meaningful. A more complete discussion of extreme percentages will appear in the technical documentation section of the NAEP web site at <http://nces.ed.gov/nationsreportcard>.

Table A.20 Average reading scale scores and standard errors, grades 4, 8, and 12: 1992–2002

	Accommodations not permitted				Accommodations permitted		
	1992	1994	1998	2000	1998	2000	2002
Grade 4	217 (0.9)	214 (1.0) *	217 (0.8)	217 (0.8) *	215 (1.1) *	213 (1.3) *	219 (0.4)
Grade 8	260 (0.9) *	260 (0.8) *	264 (0.8)	—	263 (0.8)	—	264 (0.4)
Grade 12	292 (0.6) *	287 (0.7)	291 (0.7) *	—	290 (0.6) *	—	287 (0.7)

— Data were not collected at grades 8 and 12 in 2000.

* Significantly different from 2002.

NOTE: Standard errors of the estimated scale scores appear in parentheses.

In addition to allowing for accommodations, the accommodation-permitted results at grade 4 (1998–2000) differ slightly from previous years, and from previous reported results for 1998 and 2000, due to changes in sample weighting procedures.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, 2000, and 2002 Reading Assessments.

Table A.21 Percentage of students and standard errors, by reading achievement level, grades 4, 8, and 12: 1992–2002

		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 4							
Accommodations not permitted	1992	38 (1.1)	34 (0.9)	22 (0.9) *	6 (0.6)	62 (1.1)	29 (1.2) *
	1994	40 (1.0) *	31 (0.7) *	22 (0.8) *	7 (0.7)	60 (1.0) *	30 (1.1)
	1998	38 (0.9)	32 (0.7)	24 (0.7)	7 (0.5)	62 (0.9)	31 (0.9)
	2000	37 (0.8)	31 (0.9)	24 (0.8)	8 (0.5)	63 (0.8)	32 (0.9)
Accommodations permitted	1998	40 (1.2) *	30 (0.8) *	22 (0.8) *	7 (0.5)	60 (1.2) *	29 (0.9) *
	2000	41 (1.4) *	30 (1.1) *	23 (1.0)	7 (0.6)	59 (1.4) *	29 (1.1)
	2002	36 (0.5)	32 (0.3)	24 (0.3)	7 (0.2)	64 (0.5)	31 (0.4)
Grade 8							
Accommodations not permitted	1992	31 (1.0) *	40 (0.7) *	26 (1.0) *	3 (0.3)	69 (1.0) *	29 (1.1) *
	1994	30 (0.9) *	40 (0.7) *	27 (0.8) *	3 (0.3)	70 (0.9) *	30 (0.9) *
	1998	26 (0.9)	41 (0.8) *	31 (0.9)	3 (0.4)	74 (0.9)	33 (0.9)
Accommodations permitted	1998	27 (0.8) *	41 (0.9)	30 (0.9)	3 (0.3)	73 (0.8) *	32 (1.1)
	2002	25 (0.5)	43 (0.4)	30 (0.5)	3 (0.2)	75 (0.5)	33 (0.5)
Grade 12							
Accommodations not permitted	1992	20 (0.6) *	39 (0.7)	36 (0.8) *	4 (0.3)	80 (0.6) *	40 (0.8) *
	1994	25 (0.7)	38 (0.7)	32 (0.9)	4 (0.5)	75 (0.7)	36 (1.0)
	1998	23 (0.9) *	37 (0.8)	35 (1.0) *	6 (0.4) *	77 (0.9) *	40 (0.9) *
Accommodations permitted	1998	24 (0.7) *	36 (0.6)	35 (0.8) *	6 (0.4) *	76 (0.7) *	40 (0.7) *
	2002	26 (0.8)	38 (0.6)	31 (0.8)	5 (0.3)	74 (0.8)	36 (0.8)

* Significantly different from 2002.

NOTE: Standard errors of the estimated percentages appear in parentheses.

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

In addition to allowing for accommodations, the accommodation-permitted results at grade 4 (1998–2000) differ slightly from previous years, and from previous reported results for 1998 and 2000, due to changes in sample weighting procedures.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, 2000, and 2002 Reading Assessments.

Table A.22 Average reading scale scores and standard errors, by race/ethnicity and eligibility for free/reduced-price school lunch, grades 4, 8, and 12: 2002

	Eligible	Not eligible	Information not available
Grade 4			
White	215 (0.6)	233 (0.4)	234 (1.1)
Black	193 (0.5)	212 (1.0)	206 (1.9)
Hispanic	195 (1.8)	216 (1.3)	207 (3.1)
Asian/Pacific Islander	212 (3.0)	234 (1.5)	222 (3.3)
American Indian/Alaska Native	201 (2.3)	219 (2.2)	200 (6.8)
Grade 8			
White	260 (0.6)	275 (0.5)	279 (1.4)
Black	239 (0.7)	256 (1.1)	251 (2.6)
Hispanic	244 (1.1)	256 (1.5)	249 (2.3)
Asian/Pacific Islander	249 (3.4)	274 (1.5)	276 (3.6)
American Indian/Alaska Native	240 (4.8)	265 (2.1)	255 (5.2) !
Grade 12			
White	283 (2.0)	292 (0.9)	298 (1.4)
Black	260 (1.7)	272 (1.6)	273 (3.2)
Hispanic	266 (2.2)	278 (1.9)	280 (3.8)
Asian/Pacific Islander	274 (4.3)	288 (2.8)	296 (3.8) !
American Indian/Alaska Native	***	***	***

! The nature of the sample does not allow accurate determination of the variability of the statistic.

*** Quality control activities and special analysis raised concerns about the accuracy and precision of grade 12 American Indian data. As a result, they are omitted from this report.

NOTE: Standard errors of the estimated scale scores appear in parentheses.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Table A.23 Average reading scale scores and standard errors, grade 8 public schools: By state, 1998 and 2002

Grade 8	Accommodations not permitted		Accommodations permitted	
	1998	1998	1998	2002
Nation (Public) ¹	261 (0.8)	261 (0.8) *	263 (0.5)	
Alabama	255 (1.3)	255 (1.4)	253 (1.3)	
Arizona	261 (1.2) *	260 (1.1)	257 (1.3)	
Arkansas	256 (1.3) *	256 (1.3) *	260 (1.1)	
California [‡]	253 (1.7)	252 (1.6)	250 (1.8)	
Colorado	264 (1.1)	264 (1.0)	—	
Connecticut	272 (1.1) **	270 (1.0) *	267 (1.2)	
Delaware	256 (1.3) **	254 (1.3) **	267 (0.5)	
Florida	253 (1.7) **	255 (1.4) **	261 (1.6)	
Georgia	257 (1.4)	257 (1.4)	258 (1.0)	
Hawaii	250 (1.3)	249 (1.0) *	252 (0.9)	
Idaho	—	—	266 (1.1)	
Indiana	—	—	265 (1.3)	
Kansas [‡]	268 (1.2)	268 (1.4)	269 (1.3)	
Kentucky	262 (1.3)	262 (1.4)	265 (1.0)	
Louisiana	252 (1.5) *	252 (1.4) *	256 (1.5)	
Maine	273 (1.2)	271 (1.2)	270 (0.9)	
Maryland	262 (1.8)	261 (1.8)	263 (1.7)	
Massachusetts	269 (1.6)	269 (1.4)	271 (1.3)	
Michigan	—	—	265 (1.6)	
Minnesota [‡]	267 (1.3)	265 (1.4)	—	
Mississippi	251 (1.4) *	251 (1.2) *	255 (0.9)	
Missouri	263 (1.3) **	262 (1.3) **	268 (1.0)	
Montana [‡]	270 (1.1)	271 (1.3)	270 (1.0)	
Nebraska	—	—	270 (0.9)	
Nevada	257 (1.1) **	258 (1.0) **	251 (0.8)	
New Mexico	258 (1.2) *	258 (1.2) **	254 (1.0)	
New York [‡]	266 (1.6)	265 (1.5)	264 (1.5)	
North Carolina	264 (1.1)	262 (1.1)	265 (1.1)	
North Dakota [‡]	—	—	268 (0.8)	
Ohio	—	—	268 (1.6)	
Oklahoma	265 (1.3) *	265 (1.2) *	262 (0.8)	
Oregon [‡]	266 (1.4)	266 (1.5)	268 (1.3)	
Pennsylvania	—	—	265 (1.0)	
Rhode Island	262 (1.0)	264 (0.9) *	262 (0.8)	
South Carolina	255 (1.3)	255 (1.1)	258 (1.1)	
Tennessee [‡]	259 (1.3)	258 (1.2)	260 (1.4)	
Texas	262 (1.5)	261 (1.4)	262 (1.4)	
Utah	265 (1.1)	263 (1.0)	263 (1.1)	
Vermont	—	—	272 (0.9)	
Virginia	266 (1.1)	266 (1.1)	269 (1.0)	
Washington [‡]	265 (1.3)	264 (1.2) *	268 (1.2)	
West Virginia	262 (1.2)	262 (1.0)	264 (1.0)	
Wisconsin [‡]	266 (1.6)	265 (1.8)	—	
Wyoming	262 (1.3)	263 (1.3)	265 (0.7)	
Other Jurisdictions				
American Samoa	—	—	198 (1.7)	
District of Columbia	236 (2.0)	236 (2.1)	240 (0.9)	
DDESS ²	269 (3.3)	268 (4.5)	272 (1.0)	
DoDDS ³	269 (1.0) **	269 (1.0) **	273 (0.6)	
Guam	—	—	240 (1.2)	
Virgin Islands	233 (2.9) *	231 (2.1) **	241 (1.3)	

— Indicates that the jurisdiction did not participate or did not meet minimum participation guidelines for reporting.

[‡] Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2002.

* Significantly different from 2002 when only one jurisdiction or the nation is being examined.

** Significantly different from 2002 when using a multiple-comparison procedure based on all jurisdictions that participated both years.

¹ National results that are presented for assessments prior to 2002 are based on the national sample, not on aggregated state assessment samples.

² Department of Defense Domestic Dependent Elementary and Secondary Schools. ³ Department of Defense Dependents Schools (Overseas).

NOTE: Standard errors of the estimated scale scores appear in parentheses.

Comparative performance results may be affected by changes in exclusion rates for students with disabilities and limited English proficient students in the NAEP samples.

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

Table A.24 Percentages of students at or above *Proficient* and standard errors, by race/ethnicity, grade 8 public schools: By state, 1998 and 2002

Grade 8	White			Black			Hispanic		
	Accommodations not permitted	Accommodations permitted		Accommodations not permitted	Accommodations permitted		Accommodations not permitted	Accommodations permitted	
	1998	1998	2002	1998	1998	2002	1998	1998	2002
Nation (Public) ¹	38 (1.2)	37 (1.3)	39 (0.7)	11 (1.3)	11 (1.6)	13 (0.7)	14 (1.5)	13 (1.0)	14 (0.8)
Alabama	28 (1.8)	29 (2.6)	30 (1.8)	7 (1.4)	8 (1.3)	7 (0.9)	***	***	***
Arizona	37 (1.8)	35 (1.8)	32 (2.4)	10 (4.0)	12 (4.3)	12 (4.3)	12 (1.8)	12 (2.0)	11 (1.6)
Arkansas	28 (1.5) *	29 (1.7)	34 (1.8)	6 (1.8)	5 (1.8)	6 (1.8)	***	***	***
California [‡]	35 (3.0)	35 (3.0)	33 (3.1)	12 (3.2)	9 (2.5)	13 (4.3)	8 (1.3)	8 (1.4)	10 (1.4)
Colorado	37 (1.8)	36 (1.4)	—	9 (3.7) !	10 (3.7)	—	10 (1.9)	11 (2.2)	—
Connecticut	49 (1.5)	47 (1.7)	48 (1.7)	10 (2.9)	11 (2.9)	9 (1.9)	13 (3.1)	13 (4.5)	10 (2.2)
Delaware	31 (2.0) **,*	30 (2.0) **,*	42 (1.1)	10 (1.9)	9 (1.3) *	14 (1.2)	18 (6.3) !	17 (5.9)	14 (2.7)
Florida	31 (2.1)	30 (2.1)	36 (2.4)	7 (1.3) *	7 (1.3) *	14 (1.7)	15 (3.0)	17 (3.3)	20 (3.5)
Georgia	34 (2.5)	35 (2.0)	35 (1.8)	9 (1.5)	10 (1.3)	14 (1.5)	***	***	14 (4.9)
Hawaii	31 (2.8)	30 (2.6)	30 (2.6)	***	***	18 (7.9)	***	***	16 (5.3)
Idaho	—	—	35 (2.2)	—	—	***	—	—	17 (3.1)
Indiana	—	—	34 (1.6)	—	—	12 (2.6)	—	—	***
Kansas [‡]	39 (1.9)	40 (2.0)	42 (1.9)	17 (9.3)	20 (8.4)	12 (3.2)	15 (4.3)	11 (2.4)	23 (4.5)
Kentucky	31 (1.8)	32 (1.7)	33 (1.6)	9 (2.9)	11 (3.1)	14 (3.0)	***	***	***
Louisiana	26 (1.9)	25 (2.2) *	32 (2.0)	6 (1.3)	6 (1.2)	9 (1.2)	***	***	***
Maine	42 (1.8)	42 (1.8)	38 (1.1)	***	***	***	***	***	***
Maryland	41 (2.6)	41 (2.9)	44 (2.7)	11 (1.5)	10 (1.7)	13 (1.6)	27 (6.6)	23 (6.3)	24 (5.0) !
Massachusetts	41 (2.4)	43 (1.9)	47 (1.8)	13 (3.8)	12 (3.8)	12 (2.8)	12 (3.3)	12 (3.0)	16 (2.9)
Michigan	—	—	37 (1.5)	—	—	13 (3.1)	—	—	***
Minnesota [‡]	39 (1.9)	39 (1.9)	—	8 (4.5)	7 (3.4) !	—	***	***	—
Mississippi	29 (1.9)	28 (2.2)	31 (2.4)	8 (1.1)	8 (1.1)	7 (1.0)	***	***	***
Missouri	32 (1.6)	31 (1.8) *	37 (1.7)	8 (2.6)	9 (1.7)	13 (2.6)	***	***	***
Montana [‡]	40 (1.6)	42 (1.7)	40 (1.9)	***	***	***	***	***	***
Nebraska	—	—	40 (1.3)	—	—	11 (3.5)	—	—	14 (4.0)
Nevada	30 (1.5)	29 (1.7)	25 (1.6)	10 (3.0)	10 (3.4)	7 (1.9)	10 (1.8)	9 (1.6)	8 (1.6)
New Mexico	37 (2.3)	36 (1.9)	32 (2.6)	***	***	***	14 (1.6)	15 (1.5)	12 (1.2)
New York [‡]	45 (3.0)	44 (2.2)	43 (2.7)	12 (2.2)	10 (1.7)	12 (3.0)	12 (2.1)	10 (2.6)	15 (3.1)
North Carolina	40 (1.8)	39 (1.7)	42 (2.1)	13 (2.1)	12 (1.7)	11 (1.3)	***	***	18 (6.4)
North Dakota [‡]	—	—	35 (1.3)	—	—	***	—	—	***
Ohio	—	—	40 (2.2)	—	—	13 (3.5) !	—	—	***
Oklahoma	33 (2.0)	34 (2.2)	33 (1.7)	12 (3.5)	14 (2.5)	8 (2.5)	10 (4.1)	16 (4.8)	14 (4.5)
Oregon [‡]	36 (2.1)	37 (2.2)	39 (1.9)	10 (6.4) !	10 (5.6) !	—	13 (4.0)	15 (3.6)	14 (4.1)
Pennsylvania	—	—	40 (1.7)	—	—	8 (1.2)	—	—	14 (3.6) !
Rhode Island	33 (1.5)	35 (1.5)	36 (1.3)	15 (5.5)	12 (4.5)	12 (4.8)	10 (2.9)	10 (3.2)	12 (2.1)
South Carolina	30 (1.6)	30 (1.4)	35 (2.1)	8 (1.1)	9 (1.0)	9 (1.3)	***	***	***
Tennessee [‡]	31 (2.0)	32 (1.9)	33 (1.7)	6 (1.4)	7 (1.7)	11 (1.7)	***	***	***
Texas	38 (2.4)	38 (2.6)	47 (2.8)	12 (3.7)	12 (2.5)	15 (2.3)	14 (1.8)	14 (2.1)	17 (1.5)
Utah	32 (1.2)	32 (1.5)	35 (1.3)	***	***	***	23 (6.4)	20 (4.3)	9 (2.9)
Vermont	—	—	40 (1.5)	—	—	***	—	—	***
Virginia	41 (1.8)	42 (1.6)	46 (1.8)	13 (2.1)	13 (2.2)	15 (1.7)	24 (8.1)	28 (7.1)	23 (5.4)
Washington [‡]	35 (2.0)	35 (1.9)	40 (2.0)	14 (4.9)	13 (4.7)	18 (4.2)	12 (4.0)	11 (2.7)	20 (4.5)
West Virginia	28 (1.2)	28 (1.1)	30 (1.6)	11 (6.1)	11 (4.1)	10 (4.8)	***	***	***
Wisconsin [‡]	37 (2.2)	37 (1.8)	—	8 (3.0)	10 (4.4)	—	18 (4.0) !	19 (5.4) !	—
Wyoming	31 (1.7)	32 (1.6)	33 (1.2)	***	***	***	15 (3.9)	19 (4.3)	13 (3.4)
Other Jurisdictions									
American Samoa	—	—	***	—	—	***	—	—	***
District of Columbia	***	***	***	9 (1.2)	9 (1.1)	8 (0.9)	15 (7.2)	22 (6.8)	11 (3.4)
DDESS ²	45 (3.8)	48 (5.5)	48 (4.1)	21 (6.0)	20 (7.6)	19 (3.9)	37 (6.5)	43 (6.3)	37 (5.0)
DoDDS ³	45 (3.8)	45 (2.3)	48 (2.1)	24 (2.2)	22 (5.4)	24 (2.7)	26 (5.2)	27 (5.9)	29 (4.6)
Guam	—	—	***	—	—	***	—	—	***
Virgin Islands	***	***	***	9 (2.9)	8 (1.9)	7 (1.4)	***	***	4 (2.8)

See footnotes at end of table. ▶

Table A.24 Percentages of students at or above Proficient and standard errors, by race/ethnicity, grade 8 public schools: By state, 1998 and 2002 — Continued

Grade 8	Asian/Pacific Islander			American Indian/Alaska Native			Other		
	Accommodations not permitted	Accommodations permitted		Accommodations not permitted	Accommodations permitted		Accommodations not permitted	Accommodations permitted	
	1998	1998	2002	1998	1998	2002	1998	1998	2002
Nation (Public) ¹	32 (6.0)	30 (6.1)	34 (2.0)	***f***	***f***	18 (2.2)	***f***	***f***	24 (4.1)
Alabama	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***
Arizona	***f***	***f***	***f***	10 (4.1)	7 (2.4) !	12 (3.0) !	***f***	***f***	***f***
Arkansas	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***
California [‡]	24 (4.7)	25 (3.7)	25 (4.6)	***f***	***f***	***f***	***f***	***f***	***f***
Colorado	30 (6.6)	25 (7.2)	—	***f***	***f***	—	***f***	***f***	—
Connecticut	59 (7.6) *	58 (8.4)	34 (5.0)	***f***	***f***	***f***	***f***	***f***	***f***
Delaware	***f***	***f***	54 (5.4)	***f***	***f***	***f***	***f***	***f***	***f***
Florida	54 (7.0)	47 (7.6)	***f***	***f***	***f***	***f***	***f***	***f***	***f***
Georgia	***f***	***f***	27 (5.5)	***f***	***f***	***f***	***f***	***f***	***f***
Hawaii	16 (1.2)	16 (1.3)	17 (1.3)	***f***	***f***	***f***	17 (2.9)	17 (2.9)	24 (3.4)
Idaho	—	—	***f***	—	—	***f***	—	—	***f***
Indiana	—	—	***f***	—	—	***f***	—	—	***f***
Kansas [‡]	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***
Kentucky	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***
Louisiana	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***
Maine	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***
Maryland	53 (7.1)	55 (7.5)	56 (6.8)	***f***	***f***	***f***	***f***	***f***	***f***
Massachusetts	35 (7.5)	40 (6.0)	37 (7.3)	***f***	***f***	***f***	***f***	***f***	***f***
Michigan	—	—	***f***	—	—	***f***	—	—	***f***
Minnesota [‡]	21 (7.4)	16 (4.3)	—	***f***	***f***	—	***f***	***f***	—
Mississippi	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***
Missouri	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***
Montana [‡]	***f***	***f***	***f***	20 (6.2) !	20 (5.9) !	17 (3.9) !	***f***	***f***	***f***
Nebraska	—	—	***f***	—	—	***f***	—	—	***f***
Nevada	21 (5.4)	24 (4.9)	24 (4.6)	***f***	***f***	***f***	***f***	***f***	***f***
New Mexico	***f***	***f***	***f***	10 (2.9)	11 (4.0)	9 (1.9)	***f***	***f***	***f***
New York [‡]	43 (9.5) !	49 (8.4) !	36 (6.8) !	***f***	***f***	***f***	***f***	***f***	***f***
North Carolina	***f***	***f***	***f***	21 (6.0) !	21 (6.4) !	***f***	***f***	***f***	***f***
North Dakota [‡]	—	—	***f***	—	—	19 (6.0) !	—	—	***f***
Ohio	—	—	***f***	—	—	***f***	—	—	***f***
Oklahoma	***f***	***f***	***f***	22 (3.8)	23 (3.7)	23 (2.6)	***f***	***f***	***f***
Oregon [‡]	33 (6.9)	35 (7.4)	41 (5.3)	***f***	***f***	***f***	***f***	***f***	***f***
Pennsylvania	—	—	27 (7.5) !	—	—	***f***	—	—	***f***
Rhode Island	34 (6.2)	30 (6.9)	19 (4.3)	***f***	***f***	***f***	***f***	***f***	***f***
South Carolina	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***
Tennessee [‡]	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***
Texas	45 (8.5)	43 (8.1)	39 (9.2) !	***f***	***f***	***f***	***f***	***f***	***f***
Utah	***f***	***f***	22 (5.3)	***f***	***f***	***f***	***f***	***f***	***f***
Vermont	—	—	***f***	—	—	***f***	—	—	***f***
Virginia	43 (8.5)	38 (8.1)	50 (5.3)	***f***	***f***	***f***	***f***	***f***	***f***
Washington [‡]	32 (4.6)	34 (4.0)	39 (7.1)	15 (5.3)	17 (7.3)	***f***	***f***	***f***	***f***
West Virginia	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***
Wisconsin [‡]	***f***	***f***	—	***f***	***f***	—	***f***	***f***	—
Wyoming	***f***	***f***	***f***	13 (5.6) !	12 (4.5)	15 (4.1)	***f***	***f***	***f***
Other Jurisdictions									
American Samoa	—	—	1 (0.7)	—	—	***f***	—	—	***f***
District of Columbia	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***
DDESS ²	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	44 (6.8)
DoDDS ³	29 (4.1)	34 (3.7)	37 (4.3)	***f***	***f***	***f***	35 (4.4)	36 (3.8)	39 (3.0)
Guam	—	—	10 (1.2)	—	—	***f***	—	—	***f***
Virgin Islands	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***	***f***

— Indicates that the jurisdiction did not participate or did not meet minimum participation guidelines for reporting.

[‡] Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2002.

! The nature of the sample does not allow accurate determination of the variability of the statistic.

* Significantly different from 2002 when only one jurisdiction or the nation is being examined. ** Significantly different from 2002 when using a multiple-comparison procedure based on all jurisdictions that participated both years.

f Sample size is insufficient to permit a reliable estimate.

¹ National results that are presented for assessments prior to 2002 are based on the national sample, not on aggregated state assessment samples.

² Department of Defense Domestic Dependent Elementary and Secondary Schools. ³ Department of Defense Dependents Schools (Overseas).

NOTE: Comparisons between the accommodations-not-permitted and accommodations-permitted results should be interpreted with caution.

Standard errors of the estimated percentages appear in parentheses.

Comparative performance results may be affected by changes in exclusion rates for students with disabilities and limited English proficient students in the NAEP samples.

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

Analyzing Group Differences in Averages and Percentages

Statistical tests determine whether the evidence, based on the data from the groups in the sample, is strong enough to conclude that the averages or percentages are actually different for those groups in the population. If the evidence is strong (i.e., the difference is statistically significant), the report describes the group averages or percentages as being different (e.g., one group performed higher 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.

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 as 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 0.05 level.

The following example of comparing groups addresses the problem of determining whether the average reading 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 two 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 claim that group A outperformed 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 1998 and 2002 results for a particular state or subgroup) or when comparing state results with each other. 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 2002 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, 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 is¹⁹:

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

where p is the proportion of the total group contained in the subgroup. This formula was used for this report when a state was compared to the aggregate nation or a school district was compared to the entire state it belongs to.

Conducting Multiple Tests

The procedures in the previous section and the certainty ascribed to intervals (e.g., a 95 percent confidence interval) are based on statistical theory that assumes that only one confidence interval or test of statistical significance is being performed. However, 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., 0.05), adjustments (called “multiple comparison procedures”)²⁰ must be made to the methods described in the previous section. One such procedure, the Benjamini-Hochberg False Discovery Rate (FDR) procedure was used to control the certainty level.²¹

¹⁹ 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.

²⁰ Miller, R. G. (1981). *Simultaneous Statistical Inference* (2nd ed.). New York: Springer-Verlang.

²¹ 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*, no. 1, 289–300.

Unlike the 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. Furthermore, the FDR procedure used in NAEP is considered appropriately less conservative than familywise procedures for large families of comparisons.²² Therefore, the FDR procedure is more suitable for multiple comparisons in NAEP than other procedures. A detailed description of the FDR procedure will appear in the technical documentation

section of the NAEP web site at <http://nces.ed.gov/nationsreportcard>.

To illustrate how the FDR procedure is used, consider the comparisons of current and previous years' average reading scale scores for the five groups presented in table A.25. Note that the difference in average scale scores and the estimated standard error of the difference are calculated in a way comparable with that of 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.25 Example of False Discovery Rate comparisons of average scale scores for different groups of students

	Previous year		Current year		Previous year and current year			
	Average scale score	Standard error	Average scale score	Standard error	Difference in averages	Standard error of difference	Test statistic	Percent confidence ¹
Group 1	224	1.3	226	1.0	2.08	1.62	1.29	20
Group 2	187	1.7	193	1.7	6.31	2.36	2.68	1
Group 3	191	2.6	197	1.7	6.63	3.08	2.15	4
Group 4	229	4.4	232	4.6	3.24	6.35	0.51	62
Group 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.

²² 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.

The difference in average scale scores and its estimated standard error can be used to find an approximately 95 percent confidence interval as in the example in the previous section or they can be used to identify a confidence percentage. In the example in the previous section, because an approximately 95 percent confidence interval was desired, the number 1.96 was used to multiply the estimated standard error of the difference to create the approximate confidence interval. In the current example, the confidence interval for the test statistics is identified from statistical tables. Instead of checking to see if zero is within the 95 percent confidence interval about the mean, the significance level from the statistical tables can be directly compared to $100 - 95 = 5$ percent.

If the comparison of average scale scores across two years was made for only one of the five groups, there would be a significant difference between the average scale scores for the two years if the significance level were less than 5 percent. However, because we are interested 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 to us, 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,²³ 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 in the current year and previous years' average scale scores for the group 2 students is significant; for all of the other groups, average scale scores for 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. NCES is continuing to evaluate the use of FDR and multiple-comparison procedures for future reporting.

²³ 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.

NAEP Reporting Groups

Results are provided for groups of students defined by shared characteristics—gender, race or ethnicity, school’s type of location, Title I participation, eligibility for free/reduced-price school lunch, and type of school. Based on participation rate criteria, results are reported for subpopulations only when sufficient numbers of students and adequate school representation are present. The minimum requirement is at least 62 students in a particular subgroup from at least five primary sampling units (PSUs).²⁴ However, the data for all students, regardless of whether their subgroup was reported separately, were included in computing overall results. Definitions of the subpopulations are presented below.

Gender

Results are reported separately for males and females.

Race/Ethnicity

In all NAEP assessments, data about student race/ethnicity is collected from two sources: school records and student self-reports. Previously, NAEP has used student self-reported race as the primary race/ethnicity reporting variable. In 2002, it was decided to change the student race/ethnicity variable highlighted in NAEP reports. Starting in 2002, school-recorded race will become the race/ethnicity variable presented in NAEP reports. The mutually exclusive racial/ethnic categories were White, Black, Hispanic, Asian/Pacific Islander, American Indian (including Alaska Native), and Other. Information

based on student self-reported race/ethnicity will continue to be available on the NAEP Data Tool (<http://nces.ed.gov/nationsreportcard/naepdata>).

Type of Location

Results from the 2002 assessment are reported for students attending schools in three mutually exclusive location types:

Central city: This category includes central cities of all Consolidated Metropolitan Statistical Area (CMSA) or Metropolitan Statistical Area (MSA) as defined by the Office of Management and Budget. Central city is a geographical term and is not synonymous with “inner city.”

Urban fringe/large town: The urban fringe category includes any incorporated place, census designated place, or non-place territory within a CMSA or MSA of a large or mid-sized city and defined as urban by the U.S. Census Bureau, but which do not qualify as central city. A large town is defined as a place outside a CMSA or MSA with a population greater than or equal to 25,000.

Rural/small town: Rural includes all places and areas with populations of less than 2,500 that are classified as rural by the U.S. Census Bureau. A small town is defined as a place outside a CMSA or MSA with a population of less than 25,000, but greater than or equal to 2,500.

Results for each type of location are not compared across years. This is due to new methods used by NCES to identify the type of location assigned to each school in the Common Core of Data (CCD). The new methods were put into place by NCES in

²⁴ For the NAEP national assessments prior to 2002, a PSU is a selected geographic region (a county, group of counties, or metropolitan statistical area). In 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 technical documentation section of the NAEP web site at <http://nces.ed.gov/nationsreportcard>.

order to improve the quality of the assignments, and they take into account more information about the exact physical location of the school. The variable was revised in NAEP beginning with the 2000 assessments.

Title I Participation

Based on available school records, students were classified either as currently participating in a Title I program, receiving Title I services, or as not receiving such services. The classification applies only to the school year when the assessment was administered (i.e., the 2001–02 school year) and is not based on participation in previous years. If the school does not offer any Title I programs or services, all students in that school would be classified as not participating.

Eligibility for Free/Reduced-Price School Lunch

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, and reduced-price lunch qualification is set at 170 percent of the poverty level. The classification applies

only to the school year when the assessment was administered (i.e., the 2001–02 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.”

Type of School

Results are reported by the type of school that the student attends—public or nonpublic. Nonpublic schools include Catholic and other private schools.²⁵ Because they are funded by federal authorities, not state/local governments, Bureau of Indian Affairs (BIA) schools and Department of Defense Domestic Dependent Elementary and Secondary Schools (DDESS) are not included in either the public or nonpublic categories; they are included in the overall national results.

Grade 12 Participation Rates

NAEP has been described as a “low-stakes” assessment. That is, students receive no individual scores, and their NAEP performance has no effect on their grades, promotions, or graduation. There has been continued concern that this lack of consequences affects participation rates of students and schools, as well as the motivation of students to perform well on NAEP. Of particular concern has been the performance of twelfth-graders, who typically have lower student participation rates than fourth- and eighth-graders and who are more likely to omit responses compared to their younger cohorts.

²⁵ A more detailed breakdown of nonpublic school results are available on the NAEP web site (<http://nces.ed.gov/nationsreportcard/naepdata>).

In NAEP, there has been a consistent pattern of lower participation rates for older students. In the 2002 NAEP assessments, for example, the student participation rates were 94 percent and 92 percent at grades 4 and 8 respectively. At grade 12, however, the participation rate was 74 percent. School participation rates (the percentage of sampled schools that participated in the assessment) have also typically decreased with grade level. In the 2002 assessments, the national school participation rate was 85 percent for the fourth grade, 83 percent for the eighth grade, and 75 percent for the twelfth grade.

The effect of participation rates on student performance, however, is unclear. Students may choose not to participate in NAEP for many reasons such as desire to attend regular classes and not miss important instruction or conflict with other school-based activities. Similarly, there are a variety of reasons for which various schools do not participate. The sampling weights and nonresponse adjustments, described earlier in this document, provide an approximate statistical adjustment for nonparticipation. However, the effect of some school and student nonparticipation may have some undetermined effect on results.

More research is needed to delineate the factors that contribute to nonparticipation and lack of motivation. To that end, NCES is currently investigating how various types of incentives can be effectively used to increase participation in NAEP. One report that examines the impact of monetary incentives on student effort and performance is available on the NCES web site at <http://nces.ed.gov/pubsearch> (enter NCES# 2001024).

Cautions in Interpretations

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

A caution is also warranted for some small population group estimates. At times in this report, smaller population groups show very large increases or decreases across years in average scores. For example, fourth-grade Hispanic students in Delaware are reported as having a 36-point score increase between 1998 and 2002. However, it is often necessary to interpret such score gains with extreme caution. For one thing, the effects of exclusion-rate changes for small subgroups may be more marked for small groups than they are for the whole population. To continue with the Delaware example, 2 percent of Hispanic students were excluded in 1998. This number increased to 21 percent in 2002. Also, the standard errors are often quite large around the score estimates for small groups, which in turn means the standard error around the gain is also large. While the Delaware Hispanic student scores went up 36 points, the standard error of the gain is almost 12 points.

B

Appendix B Subgroup Percentage Appendix

Table B.1 Weighted percentage of students, by gender, grades 4, 8, and 12: 1992–2002

		Accommodations not permitted				Accommodations permitted		
		1992	1994	1998	2000	1998	2000	2002
Grade 4								
	Male	51	51	50	50	50	50	51
	Female	49	49	50	50	50	50	49
Grade 8								
	Male	51	50	50	—	51	—	50
	Female	49	50	50	—	49	—	50
Grade 12								
	Male	49	50	48	—	49	—	49
	Female	51	50	52	—	51	—	51

— Data were not collected at grades 8 and 12 in 2000.

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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, 2000, and 2002 Reading Assessments.

Table B.2 Weighted percentage of students, by race/ethnicity, grades 4, 8, and 12: 1992–2002

	Accommodations not permitted				Accommodations permitted		
	1992	1994	1998	2000	1998	2000	2002
Grade 4							
White	73	72	70	69	66	63	61
Black	17	17	16	16	15	17	17
Hispanic	7	7	10	11	14	14	16
Asian/Pacific Islander	2	3	3	3	4	4	4
American Indian/Alaska Native	1	1	1	1	1	1	1
Other	#	#	#	#	1	1	1
Grade 8							
White	72	72	70	—	70	—	65
Black	16	16	15	—	15	—	15
Hispanic	8	8	11	—	11	—	14
Asian/Pacific Islander	3	3	3	—	3	—	4
American Indian/Alaska Native	1	1	#	—	#	—	1
Other	1	#	#	—	#	—	1
Grade 12							
White	74	75	72	—	72	—	71
Black	15	13	14	—	14	—	12
Hispanic	7	7	10	—	10	—	10
Asian/Pacific Islander	3	4	4	—	4	—	5
American Indian/Alaska Native	#	1	#	—	#	—	#
Other	1	#	#	—	#	—	1

— Data were not collected at grades 8 and 12 in 2000.

Percentage rounds to zero.

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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, 2000, and 2002 Reading Assessments.

Table B.3 Weighted percentage of students, by eligibility for free/reduced-price school lunch, grades 4, 8, and 12: 1998–2002

	Accommodations not permitted		Accommodations permitted		
	1998	2000	1998	2000	2002
Grade 4					
Eligible	35	34	38	38	40
Not eligible	54	51	51	48	47
Information not available	12	15	11	14	13
Grade 8					
Eligible	27	—	28	—	31
Not eligible	56	—	56	—	54
Information not available	17	—	17	—	15
Grade 12					
Eligible	14	—	14	—	19
Not eligible	67	—	67	—	64
Information not available	19	—	19	—	17

— Data were not collected at grades 8 and 12 in 2000.

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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998, 2000, and 2002 Reading Assessments.

Table B.4 Weighted percentage of students, by eligibility for free/reduced-price school lunch and race/ethnicity, grades 4, 8, and 12: 2002

	Eligible	Not eligible	Information not available
Grade 4			
White	24	62	14
Black	68	24	8
Hispanic	68	19	13
Asian/Pacific Islander	33	47	20
American Indian/Alaska Native	59	33	8
Grade 8			
White	19	65	16
Black	58	31	11
Hispanic	58	28	15
Asian/Pacific Islander	31	47	21
American Indian/Alaska Native	55	33	12
Grade 12			
White	11	70	19
Black	39	48	12
Hispanic	42	41	17
Asian/Pacific Islander	24	64	12
American Indian/Alaska Native	***	***	***

*** Quality control activities and special analysis raised concerns about the accuracy and precision of grade 12 American Indian data. As a result, they are omitted from this report.

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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Table B.5 Weighted percentage of students, by school participation in Title I, grades 4, 8, and 12: 2002

		2002
Grade 4		
	Participated	33
	Did not participate	67
Grade 8		
	Participated	19
	Did not participate	81
Grade 12		
	Participated	10
	Did not participate	90

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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Table B.6 Weighted percentage of students, by student-reported parents' highest level of education, grades 8 and 12: 1992–2002

	Accommodations not permitted			Accommodations permitted	
	1992	1994	1998	1998	2002
Grade 8					
Less than high school	8	7	7	7	7
Graduated high school	24	22	22	22	17
Some education after high school	19	20	18	18	19
Graduated college	41	43	44	44	48
Unknown	8	9	9	9	9
Grade 12					
Less than high school	8	7	7	7	7
Graduated high school	22	21	19	19	18
Some education after high school	27	26	25	25	24
Graduated college	41	44	46	46	48
Unknown	2	3	3	3	3

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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, and 2002 Reading Assessments.

Table B.7 Weighted percentage of students, by type of school, grades 4, 8, and 12: 1992–2002

	Accommodations not permitted				Accommodations permitted		
	1992	1994	1998	2000	1998	2000	2002
Grade 4							
Public	89	90	89	89	90	90	90
Nonpublic	11	10	11	11	10	10	10
Nonpublic: Catholic	8	7	7	6	6	6	6
Nonpublic: Other	4	4	4	5	4	5	5
Grade 8							
Public	89	89	89	—	89	—	91
Nonpublic	11	11	11	—	11	—	9
Nonpublic: Catholic	6	7	7	—	7	—	5
Nonpublic: Other	4	4	4	—	4	—	4
Grade 12							
Public	87	90	89	—	89	—	91
Nonpublic	13	10	11	—	11	—	9
Nonpublic: Catholic	9	6	8	—	8	—	5
Nonpublic: Other	4	4	4	—	4	—	4

— Data were not collected at grades 8 and 12 in 2000.

NOTE: Percentages may not add to 100, or to the exact nonpublic percentages, due to rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, 2000, and 2002 Reading Assessments.

Table B.8 Weighted percentage of students, by parents' highest level of education and type of school, grades 8 and 12: 2002

	Less than high school	Graduated high school	Some education after high school	Graduated college	Unknown
	Grade 8				
Public	7	18	20	46	9
Nonpublic	2	10	15	68	5
Grade 12					
Public	7	19	25	46	3
Nonpublic	2	11	19	67	1

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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2002 Reading Assessment.

Table B.9 Weighted percentage of students, by type of location, grades 4, 8, and 12: 2000 and 2002

	Accommodations not permitted	Accommodations permitted	
	2000	2000	2002
Grade 4			
Central city	32	33	30
Urban fringe/large town	45	45	42
Rural/small town	23	23	28
Grade 8			
Central city	—	—	29
Urban fringe/large town	—	—	42
Rural/small town	—	—	29
Grade 12			
Central city	—	—	28
Urban fringe/large town	—	—	41
Rural/small town	—	—	31

— Data were not collected at grades 8 and 12 in 2000.

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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2000 and 2002 Reading Assessments.

Table B.10 Weighted percentage of students, by gender, grade 4: By state, 1992–2002

Grade 4	Male					Female				
	Accommodations not permitted			Accommodations permitted		Accommodations not permitted			Accommodations permitted	
	1992	1994	1998	1998	2002	1992	1994	1998	1998	2002
Nation (Public)	51	51	50	50	51	49	49	50	50	49
Alabama	52	51	51	51	49	48	49	49	49	51
Arizona	48	50	49	49	51	52	50	51	51	49
Arkansas	50	50	50	51	53	50	50	50	49	47
California †	49	51	48	47	53	51	49	52	53	47
Colorado	51	50	49	50	—	49	50	51	50	—
Connecticut	51	50	47	49	52	49	50	53	51	48
Delaware	50	49	51	51	49	50	51	49	49	51
Florida	51	49	50	50	50	49	51	50	50	50
Georgia	51	48	50	50	51	49	52	50	50	49
Hawaii	51	51	50	50	51	49	49	50	50	49
Idaho	50	—	—	—	53	50	—	—	—	47
Indiana	50	49	—	—	50	50	51	—	—	50
Iowa †	50	51	50	51	50	50	49	50	49	50
Kansas †	—	—	53	53	50	—	—	47	47	50
Kentucky	53	51	50	50	52	47	49	50	50	48
Louisiana	50	49	49	50	51	50	51	51	50	49
Maine	48	50	51	52	53	52	50	49	48	47
Maryland	49	52	49	50	52	51	48	51	50	48
Massachusetts	50	50	48	48	51	50	50	52	52	49
Michigan	50	—	49	49	51	50	—	51	51	49
Minnesota †	51	51	51	51	52	49	49	49	49	48
Mississippi	52	49	49	49	52	48	51	51	51	48
Missouri	50	51	52	51	50	50	49	48	49	50
Montana †	—	51	50	51	51	—	49	50	49	49
Nebraska	52	51	—	—	50	48	49	—	—	50
Nevada	—	—	50	50	51	—	—	50	50	49
New Hampshire	51	50	51	51	—	49	50	49	49	—
New Jersey	50	49	—	—	—	50	51	—	—	—
New Mexico	50	48	49	50	50	50	52	51	50	50
New York †	52	50	49	48	48	48	50	51	52	52
North Carolina	51	51	49	50	49	49	49	51	50	51
North Dakota †	51	50	—	—	52	49	50	—	—	48
Ohio	50	—	—	—	50	50	—	—	—	50
Oklahoma	49	—	50	50	51	51	—	50	50	49
Oregon	—	—	49	49	50	—	—	51	51	50
Pennsylvania	48	50	—	—	53	52	50	—	—	47
Rhode Island	51	49	53	53	51	49	51	47	47	49
South Carolina	48	51	48	49	51	52	49	52	51	49
Tennessee †	50	49	50	50	52	50	51	50	50	48
Texas	52	50	50	51	48	48	50	50	49	52
Utah	48	50	52	52	51	52	50	48	48	49
Vermont	—	—	—	—	51	—	—	—	—	49
Virginia	51	50	50	50	51	49	50	50	50	49
Washington †	—	52	51	51	50	—	48	49	49	50
West Virginia	51	51	48	48	49	49	49	52	52	51
Wisconsin †	50	49	50	51	—	50	51	50	49	—
Wyoming	51	51	51	52	52	49	49	49	48	48
Other Jurisdictions										
District of Columbia	50	50	48	48	49	50	50	52	52	51
DDESS ¹	—	—	49	49	51	—	—	51	51	49
DoDDS ²	—	50	50	50	51	—	50	50	50	49
Guam	52	51	—	—	52	48	49	—	—	48
Virgin Islands	52	—	47	47	53	48	—	53	53	47

— Indicates that the jurisdiction did not participate or did not meet minimum participation guidelines for reporting.

† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2002.

¹ Department of Defense Domestic Dependent Elementary and Secondary Schools.

² Department of Defense Dependents Schools (Overseas).

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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, and 2002 Reading Assessments.

Table B.11 Weighted percentage of students, by gender, grade 8: By state, 1998 and 2002

Grade 8	Male			Female		
	Accommodations not permitted	Accommodations permitted		Accommodations not permitted	Accommodations permitted	
		1998	1998		2002	1998
Nation (Public)	51	51	50	49	49	50
Alabama	50	50	51	50	50	49
Arizona	50	51	51	50	49	49
Arkansas	51	52	50	49	48	50
California ‡	50	51	52	50	49	48
Colorado	52	52	—	48	48	—
Connecticut	51	53	50	49	47	50
Delaware	50	50	51	50	50	49
Florida	49	49	48	51	51	52
Georgia	51	51	50	49	49	50
Hawaii	50	51	50	50	49	50
Idaho	—	—	48	—	—	52
Indiana	—	—	52	—	—	48
Kansas ‡	50	51	50	50	49	50
Kentucky	51	52	50	49	48	50
Louisiana	49	50	49	51	50	51
Maine	50	50	50	50	50	50
Maryland	51	51	50	49	49	50
Massachusetts	51	51	48	49	49	52
Michigan	—	—	49	—	—	51
Minnesota ‡	51	52	—	49	48	—
Mississippi	49	48	48	51	52	52
Missouri	52	52	49	48	48	51
Montana ‡	48	48	52	52	52	48
Nebraska	—	—	53	—	—	47
Nevada	52	52	51	48	48	49
New Mexico	49	48	52	51	52	48
New York ‡	49	50	51	51	50	49
North Carolina	48	49	49	52	51	51
North Dakota ‡	—	—	52	—	—	48
Ohio	—	—	51	—	—	49
Oklahoma	50	49	50	50	51	50
Oregon ‡	51	51	49	49	49	51
Pennsylvania	—	—	50	—	—	50
Rhode Island	50	50	49	50	50	51
South Carolina	48	48	49	52	52	51
Tennessee ‡	49	49	51	51	51	49
Texas	50	50	49	50	50	51
Utah	51	51	50	49	49	50
Vermont	—	—	50	—	—	50
Virginia	50	50	50	50	50	50
Washington ‡	51	52	49	49	48	51
West Virginia	50	50	49	50	50	51
Wisconsin ‡	50	51	—	50	49	—
Wyoming	52	52	51	48	48	49
Other Jurisdictions						
American Samoa	—	—	48	—	—	52
District of Columbia	48	47	47	52	53	53
DDESS ¹	52	54	49	48	46	51
DoDDS ²	51	51	50	49	49	50
Guam	—	—	51	—	—	49
Virgin Islands	48	48	45	52	52	55

— Indicates that the jurisdiction did not participate or did not meet minimum participation guidelines for reporting.

‡ Indicates that the jurisdiction or national aggregate did not meet one or more of the guidelines for school participation in 2002.

¹ Department of Defense Domestic Dependent Elementary and Secondary Schools.

² Department of Defense Dependents Schools (Overseas).

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

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

Table B.12 Weighted percentage of students, by race/ethnicity, grade 4: By state, 1992–2002

Grade 4	White					Black					Hispanic				
	Accommodations not permitted			Accommodations permitted		Accommodations not permitted			Accommodations permitted		Accommodations not permitted			Accommodations permitted	
	1992	1994	1998	1998	2002	1992	1994	1998	1998	2002	1992	1994	1998	1998	2002
Nation (Public)	72	71	69	64	60	18	18	17	16	18	7	7	10	14	17
Alabama	65	66	65	65	60	33	32	33	33	37	#	#	1	1	1
Arizona	61	63	59	60	51	5	4	5	5	6	23	25	29	28	34
Arkansas	75	76	74	75	70	23	23	23	23	24	#	1	2	2	4
California †	51	48	47	46	34	8	7	9	9	7	28	30	29	29	47
Colorado	74	74	74	75	—	5	5	7	7	—	17	16	15	15	—
Connecticut	76	74	75	76	71	12	13	12	12	13	10	10	9	8	12
Delaware	68	68	64	62	58	27	28	29	31	33	3	2	3	5	6
Florida	63	61	55	56	49	24	24	27	27	25	11	14	15	15	22
Georgia	60	60	54	55	53	37	35	41	40	37	1	2	2	2	5
Hawaii	23	22	18	19	18	3	3	3	3	3	3	3	2	2	3
Idaho	92	—	—	—	84	#	—	—	—	1	6	—	—	—	11
Indiana	87	86	—	—	80	11	11	—	—	12	1	2	—	—	4
Iowa †	93	94	91	91	88	3	3	4	4	5	2	2	2	2	4
Kansas †	—	—	80	79	77	—	—	11	11	8	—	—	6	7	11
Kentucky	90	88	87	88	86	10	11	10	10	11	#	1	#	#	1
Louisiana	54	53	52	52	47	44	43	45	44	49	1	2	1	1	2
Maine	98	98	96	97	96	#	1	1	1	2	#	#	#	#	1
Maryland	63	61	55	55	52	31	34	35	35	36	2	2	4	4	5
Massachusetts	84	81	82	82	78	8	8	6	6	9	4	6	7	7	8
Michigan	80	—	78	78	72	15	—	17	17	21	2	—	3	3	4
Minnesota †	92	91	87	86	81	3	3	6	6	6	1	1	2	2	4
Mississippi	42	49	53	53	47	57	50	46	46	51	#	#	#	#	1
Missouri	83	81	80	80	80	15	16	16	16	17	1	1	2	2	2
Montana †	—	88	89	89	85	—	1	1	1	1	—	1	1	1	2
Nebraska	89	89	—	—	82	6	4	—	—	6	3	4	—	—	8
Nevada	—	—	66	65	54	—	—	10	10	10	—	—	17	17	27
New Hampshire	97	97	96	96	—	1	1	1	1	—	1	1	1	1	—
New Jersey	69	64	—	—	—	16	17	—	—	—	11	12	—	—	—
New Mexico	47	41	40	39	37	3	3	3	3	2	44	43	43	44	47
New York †	63	58	61	62	55	15	23	18	17	20	16	14	15	15	19
North Carolina	66	68	65	65	58	30	28	29	29	33	1	1	3	3	5
North Dakota †	96	92	—	—	87	#	1	—	—	1	#	1	—	—	1
Ohio	85	—	—	—	75	12	—	—	—	21	1	—	—	—	2
Oklahoma	78	—	70	70	62	8	—	9	9	11	3	—	6	5	7
Oregon	—	—	83	81	78	—	—	3	3	3	—	—	7	9	11
Pennsylvania	82	80	—	—	76	13	16	—	—	17	3	2	—	—	4
Rhode Island	82	83	78	79	75	6	6	7	7	8	7	6	9	9	13
South Carolina	57	57	57	56	55	41	41	41	41	42	#	1	1	1	2
Tennessee †	75	77	71	72	73	23	21	26	25	23	1	1	1	1	3
Texas	50	53	50	50	37	14	13	17	17	17	33	31	29	31	43
Utah	93	91	86	86	86	#	1	1	1	1	3	4	7	8	9
Vermont	—	—	—	—	95	—	—	—	—	2	—	—	—	—	1
Virginia	71	62	65	65	63	25	31	27	27	26	1	3	4	3	4
Washington †	—	79	78	79	76	—	5	5	4	6	—	6	6	6	7
West Virginia	96	96	95	95	95	2	3	4	4	4	#	#	#	#	#
Wisconsin †	87	87	83	82	—	7	5	10	10	—	3	4	3	4	—
Wyoming	90	90	87	88	83	1	1	1	1	2	6	6	7	7	9
Other Jurisdictions															
District of Columbia	5	5	5	6	3	91	90	84	84	88	3	4	8	8	7
DDESS ¹	—	—	47	48	39	—	—	29	29	26	—	—	13	13	14
DoDDS ²	—	51	47	47	47	—	20	19	18	16	—	10	6	6	7
Guam	10	8	—	—	1	2	2	—	—	1	1	—	—	—	#
Virgin Islands	1	—	2	2	1	87	—	84	84	84	11	—	13	13	13

See footnotes at end of table. ►

Table B.12 Weighted percentage of students, by race/ethnicity, grade 4: By state, 1992–2002 — Continued

Grade 4 — Continued	Asian/Pacific Islander					American Indian/Alaska Native					Other				
	Accommodations not permitted			Accommodations permitted		Accommodations not permitted			Accommodations permitted		Accommodations not permitted			Accommodations permitted	
	1992	1994	1998	1998	2002	1992	1994	1998	1998	2002	1992	1994	1998	1998	2002
Nation (Public)	2	3	2	4	4	1	1	1	1	1	#	#	#	#	1
Alabama	#	1	1	1	1	1	1	1	1	1	0	#	0	0	#
Arizona	1	3	2	2	2	9	6	5	6	6	#	#	#	#	#
Arkansas	1	1	#	#	1	#	#	1	#	#	#	#	#	#	#
California †	12	14	13	13	10	1	#	1	1	1	1	#	1	2	#
Colorado	2	4	3	2	—	1	1	1	1	—	1	#	#	#	—
Connecticut	2	3	2	2	3	#	#	1	1	#	#	1	1	1	#
Delaware	2	2	2	1	3	#	#	#	#	#	#	0	#	#	#
Florida	2	1	1	1	2	#	#	#	#	#	#	#	#	#	2
Georgia	1	2	2	2	2	#	0	#	#	#	1	1	1	1	1
Hawaii	62	59	64	63	63	#	1	#	#	#	8	12	12	13	12
Idaho	1	—	—	—	2	1	—	—	—	3	#	—	—	—	#
Indiana	#	1	—	—	1	0	#	—	—	1	#	#	—	—	2
Iowa †	2	1	2	2	2	#	#	#	#	1	#	#	#	#	#
Kansas †	—	—	1	2	2	—	—	1	1	1	—	—	#	#	#
Kentucky	#	1	#	#	1	#	0	0	0	#	#	#	1	1	1
Louisiana	1	2	1	2	1	#	#	1	1	1	#	0	#	#	#
Maine	1	1	1	1	1	#	#	1	#	#	#	#	#	#	#
Maryland	3	3	5	5	5	#	#	#	#	1	#	#	0	0	#
Massachusetts	4	4	4	3	4	#	#	#	#	#	#	1	#	1	1
Michigan	2	—	2	2	1	1	—	#	#	2	#	—	#	#	1
Minnesota †	3	3	3	4	4	1	2	2	2	4	#	#	#	#	1
Mississippi	#	1	#	#	1	#	#	#	#	#	#	#	#	#	#
Missouri	1	1	2	1	1	#	#	#	#	#	#	#	#	#	#
Montana †	—	1	1	1	1	—	9	8	8	11	—	#	#	#	#
Nebraska	1	2	—	—	1	1	1	—	—	3	#	#	—	—	0
Nevada	—	—	5	6	7	—	—	2	2	2	—	—	#	#	#
New Hampshire	1	1	2	2	—	#	#	#	#	—	1	#	#	#	—
New Jersey	4	6	—	—	—	#	#	—	—	—	#	1	—	—	—
New Mexico	1	2	2	2	1	4	10	11	11	13	1	1	1	1	1
New York †	4	3	5	5	4	#	1	#	#	#	1	1	1	1	1
North Carolina	1	1	1	2	2	2	2	1	1	1	#	#	1	1	2
North Dakota †	#	1	—	—	1	3	4	—	—	9	#	#	—	—	#
Ohio	1	—	—	—	1	#	—	—	—	0	#	—	—	—	1
Oklahoma	1	—	1	1	1	9	—	14	14	17	1	—	1	1	3
Oregon	—	—	5	4	4	—	—	2	2	2	—	—	1	1	2
Pennsylvania	1	2	—	—	2	#	#	—	—	#	#	#	—	—	#
Rhode Island	4	3	3	3	3	#	1	1	1	#	1	1	1	1	#
South Carolina	1	1	1	1	1	#	#	#	#	#	#	#	#	#	#
Tennessee †	1	#	1	1	1	#	#	1	#	#	#	#	#	#	#
Texas	2	2	3	2	3	#	#	1	1	1	1	#	#	#	#
Utah	2	3	3	3	3	1	1	2	1	1	#	#	1	1	#
Vermont	—	—	—	—	1	—	—	—	—	#	—	—	—	—	1
Virginia	2	4	3	3	4	#	0	1	1	1	#	#	#	#	2
Washington †	—	7	7	7	7	—	2	3	3	3	—	1	1	1	#
West Virginia	1	1	#	1	#	#	0	#	#	#	#	#	1	1	#
Wisconsin †	2	3	2	2	—	1	1	1	1	—	#	#	#	#	—
Wyoming	1	1	1	1	1	2	2	3	3	4	#	#	#	#	1
Other Jurisdictions															
District of Columbia	1	1	2	2	1	0	#	#	#	0	#	#	1	1	#
DDESS ¹	—	—	2	2	3	—	—	1	1	1	—	—	8	8	18
DoDDS ²	—	9	9	9	7	—	1	1	1	1	—	8	18	19	22
Guam	85	84	—	—	98	#	#	—	—	#	2	4	—	—	#
Virgin Islands	#	—	#	#	0	0	—	#	#	#	#	—	1	1	1

— Indicates that the jurisdiction did not participate or did not meet minimum participation guidelines for reporting.

Percentage rounds to zero.

† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2002.

¹ Department of Defense Domestic Dependent Elementary and Secondary Schools.

² Department of Defense Dependents Schools (Overseas).

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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, and 2002 Reading Assessments.

Table B.13 Weighted percentage of students, by race/ethnicity, grade 8: By state, 1998 and 2002

Grade 8	White			Black			Hispanic		
	Accommodations not permitted	Accommodations permitted		Accommodations not permitted	Accommodations permitted		Accommodations not permitted	Accommodations permitted	
		1998	1998		2002	1998		1998	2002
Nation (Public)	68	68	64	15	16	15	12	12	15
Alabama	64	63	61	33	34	37	1	1	1
Arizona	61	62	56	4	4	4	26	26	31
Arkansas	76	75	75	22	22	21	2	2	2
California †	42	40	35	8	9	7	37	37	45
Colorado	72	73	—	5	4	—	18	19	—
Connecticut	76	77	70	12	12	13	8	8	12
Delaware	65	64	63	28	30	29	4	3	5
Florida	57	57	58	27	27	21	13	13	17
Georgia	58	58	54	36	36	38	3	2	4
Hawaii	19	19	16	2	2	2	2	2	3
Idaho	—	—	89	—	—	1	—	—	8
Indiana	—	—	86	—	—	10	—	—	2
Kansas †	84	83	82	8	8	8	5	6	7
Kentucky	89	89	90	10	9	8	#	#	#
Louisiana	58	58	55	41	41	41	1	1	2
Maine	97	97	96	1	1	1	#	#	#
Maryland	59	59	55	32	33	35	4	3	6
Massachusetts	79	79	73	7	7	9	9	9	11
Michigan	—	—	77	—	—	18	—	—	2
Minnesota †	87	85	—	3	4	—	2	2	—
Mississippi	51	51	53	47	48	45	#	#	1
Missouri	85	85	81	13	13	16	1	1	2
Montana †	91	90	87	#	#	#	1	2	2
Nebraska	—	—	86	—	—	6	—	—	6
Nevada	68	68	60	8	8	10	17	18	22
New Mexico	42	42	38	3	3	2	45	44	45
New York †	61	60	57	18	19	20	15	15	17
North Carolina	65	64	64	28	29	29	2	1	3
North Dakota †	—	—	94	—	—	1	—	—	1
Ohio	—	—	81	—	—	15	—	—	2
Oklahoma	72	72	62	9	9	10	4	4	7
Oregon †	85	86	82	3	3	2	6	6	8
Pennsylvania	—	—	81	—	—	13	—	—	3
Rhode Island	83	82	76	6	7	7	8	7	13
South Carolina	58	58	56	40	40	41	1	1	1
Tennessee †	76	76	77	22	22	21	1	1	1
Texas	50	50	44	13	12	12	32	33	40
Utah	90	90	86	1	1	1	5	5	8
Vermont	—	—	96	—	—	1	—	—	#
Virginia	67	66	66	26	27	25	3	3	4
Washington †	80	79	78	3	4	4	7	7	6
West Virginia	96	95	95	3	3	4	#	#	#
Wisconsin †	84	85	—	9	9	—	3	3	—
Wyoming	89	89	88	1	1	1	6	6	6
Other Jurisdictions									
American Samoa	—	—	#	—	—	0	—	—	0
District of Columbia	3	3	3	87	90	88	8	6	7
DDESS ¹	42	42	41	27	30	25	23	20	19
DoDDS ²	48	48	47	19	19	17	7	7	7
Guam	—	—	1	—	—	#	—	—	#
Virgin Islands	#	#	1	90	90	83	9	9	12

See footnotes at end of table. ►

Table B.13 Weighted percentage of students, by race/ethnicity, grade 8: By state, 1998 and 2002—Continued

Grade 8—Continued	Asian/Pacific Islander			American Indian/Alaska Native			Other		
	Accommodations not permitted	Accommodations permitted		Accommodations not permitted	Accommodations permitted		Accommodations not permitted	Accommodations permitted	
		1998	1998		2002	1998		1998	2002
Nation (Public)	3	4	4	#	#	1	#	#	1
Alabama	1	1	#	#	#	#	#	#	#
Arizona	2	2	2	6	6	6	#	#	#
Arkansas	1	1	1	#	#	1	#	#	#
California †	11	11	12	1	2	1	1	1	1
Colorado	3	3	—	1	1	—	#	#	—
Connecticut	3	3	4	#	#	1	1	1	1
Delaware	2	2	2	#	#	#	#	#	0
Florida	2	3	2	#	#	#	#	#	1
Georgia	2	3	3	#	#	#	1	1	1
Hawaii	66	66	68	#	#	#	10	11	11
Idaho	—	—	1	—	—	2	—	—	#
Indiana	—	—	1	—	—	#	—	—	1
Kansas ‡	2	2	2	1	1	1	#	#	0
Kentucky	1	1	1	#	#	#	#	1	1
Louisiana	1	1	1	#	#	1	#	#	0
Maine	1	1	1	1	1	#	#	#	#
Maryland	4	4	5	#	#	#	0	0	0
Massachusetts	5	4	5	#	#	#	#	#	1
Michigan	—	—	2	—	—	1	—	—	#
Minnesota †	4	6	—	2	3	—	#	#	—
Mississippi	1	1	1	#	#	#	#	#	#
Missouri	1	1	1	#	#	#	#	#	#
Montana †	1	1	1	6	6	9	1	1	#
Nebraska	—	—	2	—	—	1	—	—	#
Nevada	4	4	7	2	2	2	0	0	0
New Mexico	1	1	1	8	8	13	1	1	1
New York †	4	4	6	#	#	#	1	1	#
North Carolina	1	1	1	4	3	1	1	1	1
North Dakota †	—	—	1	—	—	4	—	—	0
Ohio	—	—	1	—	—	#	—	—	1
Oklahoma	1	1	2	13	13	18	1	1	1
Oregon †	4	4	5	1	1	2	1	1	1
Pennsylvania	—	—	3	—	—	#	—	—	#
Rhode Island	3	3	4	#	#	#	#	#	#
South Carolina	1	1	1	#	#	#	0	0	#
Tennessee †	1	1	1	#	#	#	#	#	#
Texas	3	3	4	1	2	#	#	#	#
Utah	3	2	3	2	2	2	#	#	#
Vermont	—	—	2	—	—	1	—	—	0
Virginia	3	3	4	1	#	1	#	#	1
Washington †	7	6	9	3	3	2	#	#	#
West Virginia	#	1	1	#	#	#	#	#	0
Wisconsin †	2	2	—	1	1	—	#	#	—
Wyoming	1	1	1	3	4	3	#	#	#
Other Jurisdictions									
American Samoa	—	—	100	—	—	0	—	—	0
District of Columbia	2	1	2	#	#	0	0	0	#
DDESS ¹	1	1	4	1	1	1	7	6	10
DoDDS ²	9	9	9	1	1	1	17	16	19
Guam	—	—	98	—	—	0	—	—	1
Virgin Islands	0	0	#	#	#	#	1	1	4

— Indicates that the jurisdiction did not participate or did not meet minimum participation guidelines for reporting.

Percentage rounds to zero.

† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2002.

¹ Department of Defense Domestic Dependent Elementary and Secondary Schools.

² Department of Defense Dependents Schools (Overseas).

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

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 1998, and 2002 Reading Assessments.

Table B.14 Weighted percentage of students, by eligibility for free/reduced-price school lunch, grade 4: By state, 1998 and 2002

Grade 4	Eligible			Not eligible			Information not available		
	Accommodations not permitted	Accommodations permitted		Accommodations not permitted	Accommodations permitted		Accommodations not permitted	Accommodations permitted	
	1998	1998	2002	1998	1998	2002	1998	1998	2002
Nation (Public)	38	41	43	54	51	50	7	7	7
Alabama	49	48	55	48	49	32	3	3	13
Arizona	41	39	45	45	45	37	14	16	18
Arkansas	47	47	55	49	49	42	4	4	3
California †	42	44	46	43	43	37	15	13	16
Colorado	27	27	—	71	70	—	2	2	—
Connecticut	24	23	28	66	66	66	10	11	6
Delaware	36	39	38	62	60	59	2	1	2
Florida	48	47	56	47	49	42	4	4	2
Georgia	49	48	46	44	45	51	6	7	3
Hawaii	46	46	47	53	53	51	1	1	1
Idaho	—	—	45	—	—	47	—	—	9
Indiana	—	—	35	—	—	58	—	—	7
Iowa †	27	28	31	69	69	69	3	3	#
Kansas †	34	34	42	62	61	58	4	5	#
Kentucky	47	46	49	52	53	49	1	1	2
Louisiana	61	61	59	34	34	32	5	5	9
Maine	35	35	33	63	63	61	2	2	6
Maryland	33	33	39	65	64	58	2	3	3
Massachusetts	27	26	27	68	69	67	5	5	6
Michigan	34	33	38	61	62	57	6	5	5
Minnesota †	27	28	29	69	68	58	3	4	13
Mississippi	64	63	64	36	36	26	1	1	10
Missouri	37	38	42	60	60	55	3	3	3
Montana †	34	34	40	56	56	55	10	10	5
Nebraska	—	—	38	—	—	58	—	—	4
Nevada	34	33	38	62	62	56	5	5	6
New Hampshire	18	17	—	72	74	—	10	9	—
New Mexico	56	56	55	31	31	31	13	13	15
New York †	45	45	45	52	52	50	3	3	6
North Carolina	41	41	47	54	54	49	5	5	4
North Dakota †	—	—	32	—	—	66	—	—	3
Ohio	—	—	33	—	—	60	—	—	7
Oklahoma	48	47	52	47	48	45	5	5	3
Oregon	36	36	35	57	57	51	7	8	14
Pennsylvania	—	—	35	—	—	63	—	—	3
Rhode Island	37	35	33	63	65	54	#	#	12
South Carolina	46	47	52	53	52	43	1	1	5
Tennessee †	44	43	45	53	53	50	3	4	4
Texas	45	47	56	50	50	39	5	4	5
Utah	32	32	32	51	51	63	17	17	5
Vermont	—	—	29	—	—	67	—	—	5
Virginia	31	31	33	61	62	64	8	7	3
Washington †	33	33	33	64	64	58	3	3	9
West Virginia	48	49	50	50	50	47	1	1	3
Wisconsin †	24	25	—	71	69	—	5	6	—
Wyoming	34	33	42	62	62	55	4	4	4
Other Jurisdictions									
District of Columbia	79	78	78	12	13	21	9	9	1
DDESS ¹	50	50	32	48	48	36	2	2	32
DoDDS ²	9	9	10	19	19	23	72	73	67
Guam	—	—	58	—	—	41	—	—	#
Virgin Islands	95	95	100	0	0	0	5	5	#

— Indicates that the jurisdiction did not participate or did not meet minimum participation guidelines for reporting.

Percentage rounds to zero.

† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2002.

¹ Department of Defense Domestic Dependent Elementary and Secondary Schools.

² Department of Defense Dependents Schools (Overseas).

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

Comparative performance results may be affected by changes in exclusion rates for students with disabilities and limited English proficient students in the NAEP samples.

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

Table B.15 Weighted percentage of students, by eligibility for free/reduced-price school lunch, grade 8: By state, 1998 and 2002

Grade 8	Eligible			Not eligible			Information not available		
	Accommodations not permitted	Accommodations permitted		Accommodations not permitted	Accommodations permitted		Accommodations not permitted	Accommodations permitted	
	1998	1998	2002	1998	1998	2002	1998	1998	2002
Nation (Public)	30	30	34	58	58	57	12	11	10
Alabama	40	41	43	58	58	42	2	2	15
Arizona	34	32	35	53	53	52	13	14	13
Arkansas	37	38	44	59	58	55	4	4	2
California ^{1†}	37	40	36	44	42	47	19	18	17
Colorado	24	22	—	67	67	—	9	10	—
Connecticut	17	18	29	70	70	63	13	13	8
Delaware	27	26	32	61	60	67	12	15	1
Florida	39	40	42	52	50	53	9	10	5
Georgia	36	37	40	53	52	55	11	11	5
Hawaii	35	35	41	60	60	59	5	4	#
Idaho	—	—	33	—	—	58	—	—	8
Indiana	—	—	25	—	—	70	—	—	6
Kansas [‡]	33	33	29	65	65	68	2	2	3
Kentucky	40	39	40	57	58	57	3	4	3
Louisiana	48	49	48	45	44	37	7	7	15
Maine	24	25	23	68	67	70	8	8	7
Maryland	26	28	28	72	70	70	2	2	2
Massachusetts	23	23	28	73	72	69	4	5	3
Michigan	—	—	33	—	—	61	—	—	6
Minnesota [‡]	22	22	—	72	71	—	6	6	—
Mississippi	50	51	57	42	41	37	8	7	6
Missouri	27	28	29	70	69	65	3	3	6
Montana [‡]	24	24	29	66	66	68	10	10	2
Nebraska	—	—	35	—	—	63	—	—	2
Nevada	25	25	27	66	65	64	9	10	10
New Mexico	42	42	50	42	43	30	16	15	20
New York [‡]	37	38	38	48	46	55	15	15	7
North Carolina	30	31	37	63	62	53	7	7	10
North Dakota [‡]	—	—	24	—	—	74	—	—	1
Ohio	—	—	23	—	—	67	—	—	10
Oklahoma	34	34	46	57	57	49	10	9	5
Oregon [‡]	26	25	26	68	69	64	5	6	10
Pennsylvania	—	—	30	—	—	69	—	—	#
Rhode Island	28	28	23	71	72	62	#	#	16
South Carolina	40	41	45	56	56	51	4	4	4
Tennessee [‡]	30	33	34	65	64	56	4	3	10
Texas	37	37	45	60	60	48	3	3	7
Utah	21	21	25	68	69	65	11	9	10
Vermont	—	—	22	—	—	77	—	—	1
Virginia	22	23	26	71	70	70	7	6	3
Washington [‡]	23	23	21	66	66	57	10	10	21
West Virginia	39	39	41	57	57	58	4	4	1
Wisconsin [‡]	20	21	—	71	71	—	9	8	—
Wyoming	25	26	33	74	73	65	2	2	2
Other Jurisdictions									
American Samoa	—	—	100	—	—	0	—	—	0
District of Columbia	53	53	68	24	23	31	23	24	1
DDESS ²	35	37	24	65	63	56	0	0	20
DoDDS ³	4	5	7	23	22	23	73	73	71
Guam	—	—	30	—	—	69	—	—	1
Virgin Islands	74	74	99	0	0	#	26	26	1

— Indicates that the jurisdiction did not participate or did not meet minimum participation guidelines for reporting.

Percentage rounds to zero.

† Indicates that the jurisdiction did not meet one or more of the guidelines for school participation in 2002.

¹ Percentages by students' eligibility for free/reduced-price lunch in California do not include Los Angeles.

² Department of Defense Domestic Dependent Elementary and Secondary Schools.

³ Department of Defense Dependents Schools (Overseas).

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

Comparative performance results may be affected by changes in exclusion rates for students with disabilities and limited English proficient students in the NAEP samples.

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



Appendix C

State-Level Contextual Variables

To help place results from the NAEP 2002 state assessment program into context, this appendix presents selected state-level data from sources other than NAEP.

These data are taken from the *Digest of Education Statistics 2001*.

Table C.1 Population and public-school enrollment, from non-NAEP sources: By state, April 2000 and fall 1999

	Estimated resident populations: April 1, 2000		Enrollment in public elementary and secondary schools: Fall 1999		
	Total (in thousands)	5- to 17-year-olds (in thousands)	Total	Kindergarten through grade 8 ¹	Grades 9–12
Nation	281,422	53,118	46,857,321	33,488,158	13,369,163
Alabama	4,447	827	740,732	538,687	202,045
Alaska	627	143	134,391	95,601	38,790
Arizona	5,131	985	852,612	623,561	229,051
Arkansas	2,673	499	451,034	317,714	133,320
California	33,872	6,763	6,038,589	4,336,687	1,701,902
Colorado	4,301	803	708,109	506,568	201,541
Connecticut	3,406	618	553,993	403,913	150,080
Delaware	784	143	112,836	80,274	32,562
Florida	15,982	2,701	2,381,396	1,725,493	655,903
Georgia	8,186	1,574	1,422,762	1,044,030	378,732
Hawaii	1,212	218	185,860	133,250	52,610
Idaho	1,294	271	245,331	168,822	76,509
Illinois	12,419	2,369	2,027,600	1,462,234	565,366
Indiana	6,080	1,151	988,702	699,221	289,481
Iowa	2,926	545	497,301	335,919	161,382
Kansas	2,688	524	472,188	325,818	146,370
Kentucky	4,042	729	648,180	458,607	189,573
Louisiana	4,469	902	756,579	548,019	208,560
Maine	1,275	231	209,253	148,774	60,479
Maryland	5,296	1,003	846,582	607,125	239,457
Massachusetts	6,349	1,103	971,425	706,251	265,174
Michigan	9,938	1,924	1,725,617	1,244,586	481,031
Minnesota	4,919	957	854,034	580,363	273,671
Mississippi	2,845	571	500,716	365,357	135,359
Missouri	5,595	1,058	914,110	648,758	265,352
Montana	902	175	157,556	107,490	50,066
Nebraska	1,711	333	288,261	197,014	91,247
Nevada	1,998	366	325,610	239,625	85,985
New Hampshire	1,236	234	206,783	146,854	59,929
New Jersey	8,414	1,524	1,289,256	953,766	335,490
New Mexico	1,819	378	324,495	228,592	95,903
New York	18,976	3,451	2,887,776	2,033,748	854,028
North Carolina	8,049	1,425	1,275,925	934,725	341,200
North Dakota	642	121	112,751	74,968	37,783
Ohio	11,353	2,133	1,836,554	1,296,450	540,104
Oklahoma	3,451	656	627,032	446,719	180,313
Oregon	3,421	624	545,033	378,474	166,559
Pennsylvania	12,281	2,194	1,816,716	1,262,181	554,535
Rhode Island	1,048	184	156,454	113,520	42,934
South Carolina	4,012	745	666,780	483,725	183,055
South Dakota	755	152	131,037	89,590	41,447
Tennessee	5,689	1,024	916,202	664,393	251,809
Texas	20,852	4,262	3,991,783	2,895,853	1,095,930
Utah	2,233	509	480,255	329,185	151,070
Vermont	609	114	104,559	72,276	32,283
Virginia	7,079	1,276	1,133,994	817,143	316,851
Washington	5,894	1,120	1,003,714	694,750	308,964
West Virginia	1,808	301	291,811	203,475	88,336
Wisconsin	5,364	1,026	877,753	596,439	281,314
Wyoming	494	98	92,105	61,654	30,451
Other Jurisdictions					
American Samoa	—	—	15,477	11,899	3,578
District of Columbia	572	82	77,194	59,917	17,277
Guam	—	—	32,951	24,151	8,800
Virgin Islands	—	—	20,866	14,821	6,045

— Data were not available.

¹ Includes a number of prekindergarten students.

SOURCE: U.S. Department of Commerce, Bureau of the Census, *Current Population Reports*, Series P-25, No. 1095 at the national level, SF1-P12 and unpublished data; and U.S. Department of Education, National Center for Education Statistics, Common Core of Data surveys.

Table C.2 Poverty status of school-age children and children served under IDEA and Chapter 1, from non-NAEP sources: By state, 1998 and school years 1990–91 through 1999–2000

	Poverty status of 5- to 17-year-olds: 1998		Children (birth to age 21) served under IDEA and Chapter 1 of the Education Consolidation and Improvement Act, State Operated Programs	
	Number in poverty (in thousands)	Percent in poverty	Number of children: 1999–2000 school year	Percent change: 1990–91 to 1999–2000
Nation	9,167	17.8	6,195,113	30.1
Alabama	156	21.8	99,763	5.1
Alaska	13	9.0	17,495	18.7
Arizona	222	23.6	93,336	63.1
Arkansas	57	13.1	60,864	27.2
California	1,459	22.3	640,815	36.6
Colorado	93	12.5	76,948	34.8
Connecticut	82	13.4	74,722	15.7
Delaware	24	15.7	16,287	13.9
Florida	474	20.5	356,198	50.9
Georgia	377	24.7	164,374	61.2
Hawaii	32	14.5	22,964	74.4
Idaho	50	17.4	29,112	32.2
Illinois	308	12.1	291,221	21.8
Indiana	140	12.6	151,599	32.2
Iowa	73	14.2	71,970	18.6
Kansas	59	13.2	60,036	32.8
Kentucky	118	16.7	91,537	15.3
Louisiana	244	29.8	96,632	31.2
Maine	27	12.0	35,139	25.6
Maryland	66	8.1	111,711	22.4
Massachusetts	163	15.0	165,013	6.7
Michigan	311	14.8	213,404	27.8
Minnesota	130	12.6	107,942	33.4
Mississippi	108	19.3	62,359	2.3
Missouri	136	14.4	134,950	32.4
Montana	42	21.2	19,039	11.1
Nebraska	54	14.8	42,577	30.0
Nevada	49	12.8	35,703	93.6
New Hampshire	34	13.3	28,597	45.5
New Jersey	194	13.2	214,330	18.2
New Mexico	101	23.5	52,346	45.3
New York	848	28.9	434,347	41.3
North Carolina	277	21.3	173,067	40.6
North Dakota	28	17.2	13,612	8.9
Ohio	339	16.0	236,200	15.0
Oklahoma	120	19.9	83,149	26.6
Oregon	121	19.4	73,531	33.3
Pennsylvania	382	18.0	231,175	5.4
Rhode Island	36	20.5	29,895	41.8
South Carolina	129	17.6	103,153	32.6
South Dakota	13	9.2	16,246	8.4
Tennessee	156	14.5	126,732	20.8
Texas	809	20.1	493,850	40.8
Utah	55	11.8	55,389	16.0
Vermont	13	12.2	14,073	14.8
Virginia	92	7.9	161,298	41.5
Washington	118	10.8	116,235	36.1
West Virginia	65	25.7	50,314	16.6
Wisconsin	109	11.5	121,209	39.4
Wyoming	13	13.0	13,307	18.8
Other Jurisdictions				
American Samoa	—	—	703	93.7
District of Columbia	33	46.0	9,348	48.6
Guam	—	—	2,230	27.4
Virgin Islands	—	—	1,617	21.3

— Data were not available.

IDEA: Individuals with Disabilities Education Act.

SOURCE: U.S. Department of Commerce, Bureau of the Census, Decennial Census, Minority Economic Profiles, unpublished data; *Current Population Reports, Series P-60*, "Poverty in the United States, Money Income of Households, Families, and Persons in the United States, and Income, Poverty, and Valuation of Noncash Benefits, various years, and Money Income in the U.S.: 1999", P60-201; U.S. Department of Education, Office of Special Education and Rehabilitative Services, *Annual Report to Congress on the Implementation of the Individuals with Disabilities Act*, various years.

Table C.3 Expenditure per pupil, average teacher salary, and pupil/teacher ratio, in public schools, from non-NAEP sources: By state, school years 1998–99 and 2000–01, and fall 1999

	In public elementary and secondary schools		
	Expenditure per pupil: 1998–99	Estimated average annual salary of teachers: 2000–01	Pupil/teacher ratio: Fall 1999
Nation	\$6,508	\$42,898	16 ¹
Alabama	5,188	37,956	15 ¹
Alaska	8,404	46,986	17
Arizona	4,672	36,302	19
Arkansas	4,956	34,476	14
California	5,801	48,923	21 ¹
Colorado	5,923	39,284	17
Connecticut	9,318	52,100	14
Delaware	7,706	47,047	15
Florida	5,790	37,824	18
Georgia	6,092	42,216	16
Hawaii	6,081	41,980	17
Idaho	5,066	36,375	18
Illinois	6,762	48,053	16
Indiana	6,772	43,055	17
Iowa	6,243	36,479	15
Kansas	6,015	39,432	14
Kentucky	5,560	37,234	15
Louisiana	5,548	34,253	17
Maine	7,155	36,256	13
Maryland	7,326	44,997	17
Massachusetts	8,260	47,523	13
Michigan	7,432	49,975	18
Minnesota	6,791	40,577	15
Mississippi	4,565	32,957	16
Missouri	5,855	36,764	14
Montana	5,974	32,930	15
Nebraska	6,256	34,175	14
Nevada	5,587	40,172	19
New Hampshire	6,433	38,303	15
New Jersey	10,145	53,281	13
New Mexico	5,440	33,785	16
New York	9,344	50,920	14
North Carolina	5,656	41,167	16
North Dakota	5,442	30,891	14
Ohio	6,627	42,716	16
Oklahoma	5,303	34,434	15
Oregon	6,828	42,333	20
Pennsylvania	7,450	49,500	16
Rhode Island	8,294	48,474	14
South Carolina	5,656	37,327	15
South Dakota	5,259	30,265	14
Tennessee	5,123	37,074	15 ¹
Texas	5,685	38,614	15
Utah	4,210	36,049	22
Vermont	7,541	38,651	12
Virginia	6,350	40,197	14 ¹
Washington	6,110	42,101	20
West Virginia	6,677	35,764	14
Wisconsin	7,527	41,646	14
Wyoming	6,842	34,189	13
Other Jurisdictions			
American Samoa	2,283	—	19
District of Columbia	9,650	48,651	16 ¹
Guam	—	—	18
Virgin Islands	6,983	—	14

— Data were not available.

¹ Includes imputations for underreporting.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, Revenues and Expenditures for Public Elementary and Secondary Schools, Statistics of State School Systems, and Common Core of Data Surveys; National Education Association, Estimates of School Statistics and unpublished data, 2001.

D **Appendix D**

Sample Text from the NAEP 2002 Reading Assessment

This appendix contains the reading passages released from the NAEP 2002 reading assessment at each grade. To review passages and questions from previous NAEP assessments, please visit the NAEP web site at *<http://nces.ed.gov/nationsreportcard>*.

The Box in the Barn

By Barbara Eckfeld Conner

Jason heard his mom calling him. Instead of answering her, he slipped deeper into the tall weeds behind his house. He closed his eyes, thinking of what he had done.

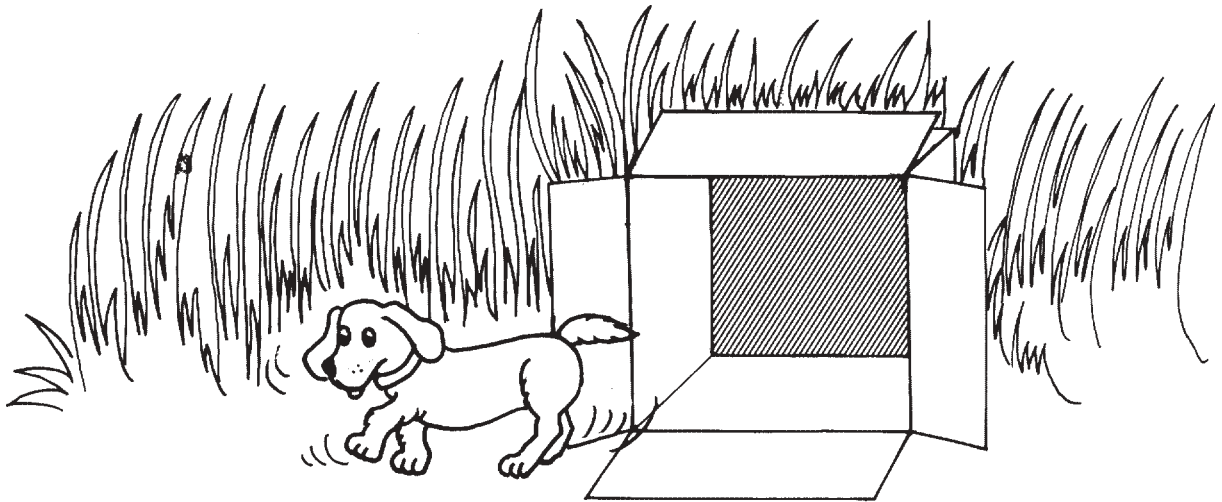
He had gotten up that morning in a good mood. Raspberry pancakes were on the table when he walked into the kitchen rubbing his eyes and yawning.

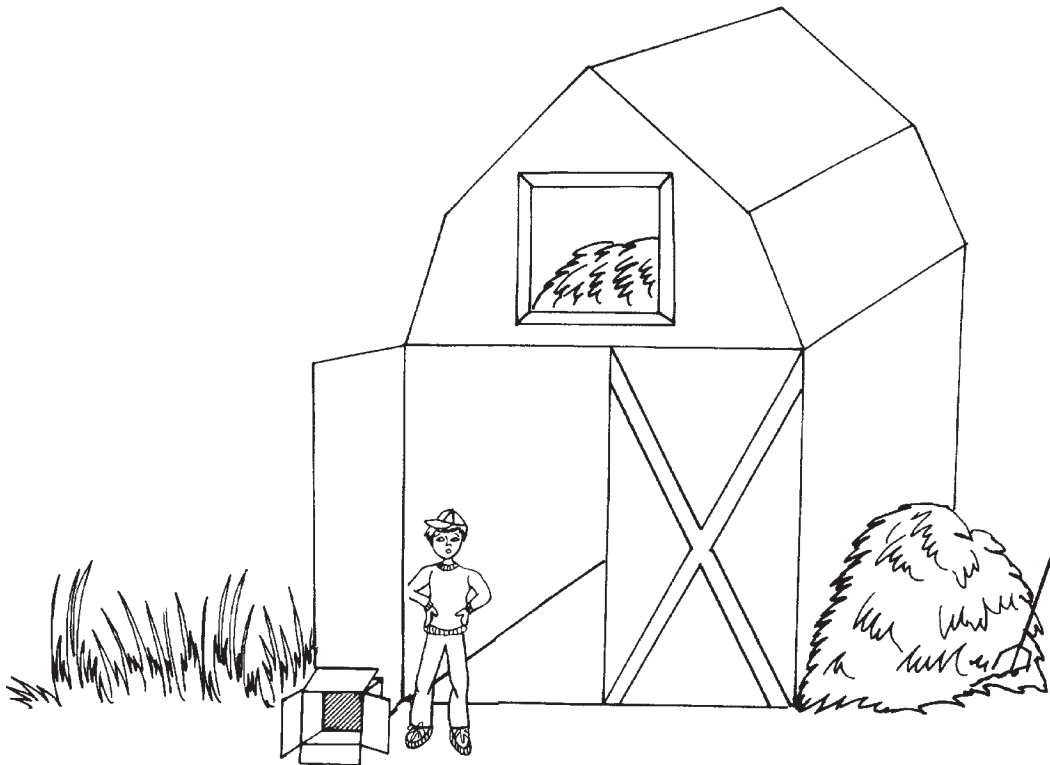
“After breakfast, Jason, I want you to go into town with me,” Mom said quietly. “It’s your sister’s birthday, and we need to shop for her gifts.”

Jason was eager to go, even if the gifts weren’t for him. Buying presents was always fun.

As they drove to town, Jason couldn’t help but ask the question that had been on his mind since yesterday when Aunt Nancy came. “What’s in the big box that Dad took to the barn, Mom? Is it something Aunt Nancy bought for Megan’s birthday?”

“It’s a surprise, Jason, and I don’t want you going near that barn today. Do you hear me?”





Jason sat staring at the road ahead. He knew that nothing would change her mind. Only now he was more curious than ever!

Back home, Megan ran out to meet Jason, her eyes wide and excited. “Jason, Jason, I’m six years old!” she cried, jumping up and down.

“I know, I know.” Jason gave her a big hug.

Soon the house was buzzing with excitement. Megan sat on the stool watching while Mom and Aunt Nancy prepared the birthday dinner. Dad wouldn’t be back for at least two hours. Jason wandered outside trying to think of something to do, but his thoughts kept returning to the box in the barn.

He started walking toward the barn, not at all sure what he’d do when he got there. He was hoping for just a glimpse of the box. Instead he heard a strange noise coming from inside the barn. He wished he could just turn back to the house, but his legs carried him into the barn. Jason saw the box. It was sitting between two bales of hay. He could hear loud wailing cries. Leaning over, Jason carefully lifted the lid. There was the most cuddly puppy he had ever seen!

“You must be pretty scared, huh, fellow?” Jason said quietly as he held the wiggly dog. “Megan’s going to love you!” He secretly wished the puppy was for him. After all, Mom and Dad knew that he had been wanting his own puppy. Probably Aunt Nancy didn’t know that, and anyway Megan would be happy.

Soon Jason was playing happily with the puppy, and he forgot that he wasn't supposed to be in the barn. Taffy, their big brown horse, stuck his head in the window as if to say, "What's going on?" Jason jumped, remembering that he wasn't supposed to be there. The puppy ran off as fast as it could out of the barn and into the field.

Jason stumbled out of the barn looking wildly for any trace of the puppy. "Come on puppy! Oh, please come here!" he called, his eyes welling up with tears.

Now here he was, two hours later, hiding in the weeds. He'd looked everywhere, but the puppy was gone. He had ruined his sister's birthday.

"Jason! It's time for dinner!" Mom called even louder now. Just when he was determined to stay forever in the tall weeds, he heard his sister's voice.

"Jason! It's time for my party, Jason!" Megan yelled excitedly.

Jason rubbed his swollen eyes, trying to look normal. He couldn't ruin everything for her. "I'm here, Megan," he called.

"Are you OK?" she asked with genuine concern.

"Sure. Let's hurry." Jason grabbed her hand as they ran back.

As soon as they reached the house, the party began. Jason tried to pretend that everything was fine. When it was time to open Megan's birthday gifts, he sat in the big easy chair, hoping no one would notice him. Finally the last present was open.

"I'll be right back," Dad said.

Jason knew Dad was going to the barn. Megan would probably never forgive him for losing her birthday puppy. Everyone, even Aunt Nancy, would be angry when they found out the puppy was gone.

"Jason! Come here!" It was Dad calling from the front yard.

Jason slowly got out of the chair. It was hard to move, but Megan grabbed his hand and said, "Come on, Jason! Let's see what Dad wants."

Jason followed Megan out the door. Mom and Aunt Nancy followed close behind.

There was Dad standing with the box next to him in the grass. "Jason, I want you to open this box and see what's inside."

Jason looked up and saw that Dad was smiling. He turned and saw that Mom, Aunt Nancy, and Megan were smiling, too. What would he say to them when there was nothing in the box? But as Jason looked down, expecting to see nothing at all, he jumped back in surprise. The puppy looked up at him, with sleepy eyes.

"Wow!" said Jason, bewildered.

"The puppy's for you, Son," his father said.

"I thought you'd like a gift, too, even if it isn't your birthday," said Aunt Nancy, laughing.

Megan started clapping. "Isn't he wonderful, Jason?" The puppy jumped up, ready to play. Jason and Megan spent the rest of the day with the puppy.

Later, when he was getting ready for bed, Jason turned to his father and said, "You know, Dad, I feel bad about something I did today."

Dad waited patiently as Jason explained what had happened. "And I still can't figure out how my puppy got back into his box!" he added.

"Well, Son, on my way home I saw your puppy running along the side of the road. I figured he had gotten out of his box somehow.... You must have felt terrible during the party," Dad continued. "I get the feeling you've learned a lot today." He pulled back the covers on Jason's bed.

Jason looked down at his new puppy, who was sleeping soundly in a basket by the bed. "Dad, I think I'll call him Buddy."

Dad smiled and tucked the covers snugly around Jason.

The Sharebots

By Carl Zimmer

When robots go to kindergarten in Maja Matarić’s lab, they learn an important lesson about how to get along in robot society.

NO MAN IS AN ISLAND, and Maja Matarić thinks no robot should be, either. Matarić, a Brandeis University computer scientist, believes robots will do their best work only when they begin to work together. “How do you get a herd of robots to do something without killing each other?” she asks. According to Matarić, you have to put them in societies and let them learn from one another, just as seagulls and baboons and people do. Matarić has already made an impressive start at teaching robots social skills. She has gotten 14 robots to cooperate at once—the biggest gaggle of machines ever to socialize.

The Nerd Herd, as Matarić calls them, are shoe-box-size machines, each of which has four wheels, two tongs to grab things, and a two-way radio. The radio allows them to triangulate their position with respect to two fixed transmitters as they wander around Matarić’s lab. It also allows them to broadcast their coordinates and other information to their neighbors. Infrared sensors help the robots find things and avoid obstacles; contact-sensitive strips tell them when they’ve crashed anyway.

Each robot is programmed with a handful of what Matarić calls behaviors—sets of instructions that enable the robot to accomplish a small goal, like following the robot in front of it. Set one robot on the floor with its wheels turned permanently to the left and program the others to follow, and they will all drive in a circle until their batteries go dead. But Matarić can get more interesting actions out of the herd by programming them to alternate among several behaviors. By telling them to home in on a target, to aggregate when they’re too far from one another, to disperse when they’re too crowded, and to avoid collisions at all times, she’s been able to get scattered robots to come together and migrate across her lab like a flock of birds.

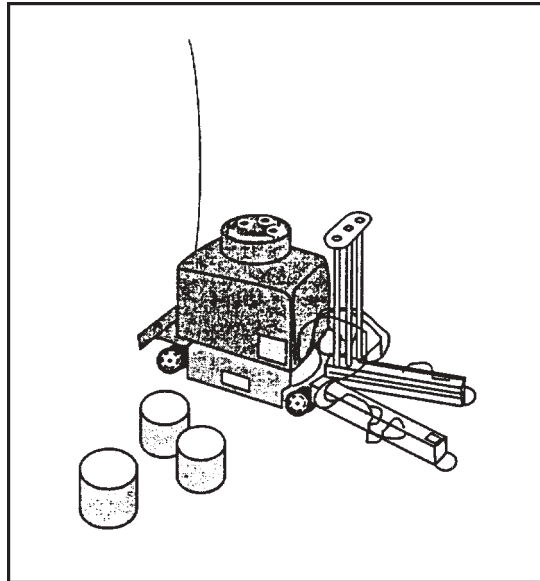
More important, the robots can also learn on their own to carry out more complex tasks. One task Matarić set for them was to forage for little metal pucks and bring them home to their nest in a corner of the lab. To give the task a natural flavor, Matarić gave the robots clocks; at “night” they had to go home and rest, and in the “morning” they looked for pucks again. In addition to five basic behaviors they could choose from, she endowed them

with a sort of prime directive: to maximize their individual point scores. Each time a robot did something right, such as locating a puck, it was automatically rewarded with points; each time it committed a blooper, such as dropping a puck, it lost points.

After some random experimentation, the robots soon learned how to forage—but not very well, because they tended to interfere with one another in their selfish pursuit of points. “Why should you ever stop and let someone else go?” asks

Matarić. “It’s always in your interest to go—but if everybody feels that way, then nobody gets through and they jam up and fight for space.” To make her creatures more efficient, though, Matarić found she didn’t have to program them with a God’s-eye view of what was good for all robots. She just

had to teach each robot to share—to let other robots know when it had found a puck, and to listen to other robots in return. “I put in the impetus to pay attention to what other robots are doing, and to try what other robots are trying, sharing the experience,” Matarić explains. “If I do something that’s good and if I say, ‘That was really great,’ then you may try it.”



**MATARIĆ'S
Nerd Herd, with
the pucks
they now pursue
collectively.**

With this simple social contract, the robots needed only 15 minutes of practice to become altruistic. They would magnanimously announce their discovery of pucks, despite having no way of knowing that this was good for the herd as a whole. At times when two robots lunged for a puck, they would stop and go through an “After you!” “No, after you!” routine, but eventually they figured out the proper way to yield. With social graces, the robot herd brought home the pucks twice as fast as without.

Matarić thinks she’ll be able to produce more complex robot societies. “I’m looking at getting specialization in the society so they can say, ‘I’ll do this, and you do that.’ If one of them has a low battery, it may become the messenger that doesn’t actually carry things. And I imagine one robot might emerge as a leader because it happens to be the most efficient. But if it stops being efficient, some other robot will take over.”

Carl Zimmer © 1995 The Walt Disney Co.
Reprinted with permission of *Discover Magazine*



Newton Minow

ADDRESS TO THE BROADCASTING INDUSTRY

I invite you to sit down in front of your television set...and keep your eyes glued to that set until the station signs off. I can assure you that you will observe a vast wasteland.

Newton Minow (1926-) was appointed by President John Kennedy as chairman of the Federal Communications Commission, the agency responsible for regulating the use of the public airwaves. On May 9, 1961, he spoke to 2,000 members of the National Association of Broadcasters and told them that the daily fare on television was “a vast wasteland.” Minow’s indictment of commercial television launched a national debate about the quality of programming. After Minow’s speech, the television critic for *The New York Times* wrote: “Tonight some broadcasters were trying to find dark explanations for Mr. Minow’s attitude. In this matter the viewer possibly can be a little helpful; Mr. Minow has been watching television.”

...**Y**our industry possesses the most powerful voice in America. It has an inescapable duty to make that voice ring with intelligence and with leadership. In a few years this exciting industry has grown from a novelty to an instrument of overwhelming impact on the American people. It should be making ready for the kind of leadership that newspapers and magazines assumed years ago, to make our people aware of their world.

Ours has been called the jet age, the atomic age, the space age. It is also, I submit, the television age. And just as history will decide whether the leaders of today’s world employed the atom to destroy the world or rebuild it for mankind’s benefit, so will history decide whether today’s broadcasters employed their powerful voice to enrich the people or debase them...

Like everybody, I wear more than one hat. I am the chairman of the FCC. I am also a television viewer and the husband and father of other television viewers. I have seen a great many television programs that seemed to me eminently worthwhile, and I am not talking about the much-bemoaned good old days of “Playhouse 90” and “Studio One.”

I am talking about this past season. Some were wonderfully entertaining, such as “The Fabulous Fifties,” the “Fred Astaire Show” and the “Bing Crosby Special”; some were dramatic and moving, such as Conrad’s “Victory” and “Twilight Zone”; some were marvelously informative, such as “The Nation’s Future,” “CBS Reports,” and “The Valiant Years.” I could list many more—programs that I am sure everyone here felt enriched his own life and that of his family. When television is good, nothing—not the theater, not the magazines or newspapers—nothing is better.

But when television is bad, nothing is worse. I invite you to sit down in front of your television set when your station goes on the air and stay there without a book, magazine, newspaper, profit-and-loss sheet, or rating book to distract you—and keep your eyes glued to that set until the station signs off. I can assure you that you will observe a vast wasteland.

You will see a procession of game shows, violence, audience participation shows, formula comedies about totally unbelievable families, blood and thunder, mayhem, violence, sadism, murder, Western badmen, Western good men, private eyes, gangsters, more violence and cartoons. And, endlessly, commercials—many

screaming, cajoling, and offending. And, most of all, boredom. True, you will see a few things you will enjoy. But they will be very, very few. And if you think I exaggerate, try it.

Is there one person in this room who claims that broadcasting can’t do better?...

Why is so much of television so bad? I have heard many answers: demands of your advertisers; competition for ever higher ratings; the need always to attract a mass audience; the high cost of television programs; the insatiable appetite for programming material—these are some of them. Unquestionably these are tough problems not susceptible to easy answers.

But I am not convinced that you have tried hard enough to solve them. I do not accept the idea that the present overall programming is aimed accurately at the public taste. The ratings tell us only that some people have their television sets turned on, and, of that number, so many are tuned to one channel and so many to another. They don’t tell us what the public might watch if they were offered half a dozen additional choices. A rating, at best, is an indication of how many people saw what you gave them. Unfortunately it does not reveal the depth of the penetration or the intensity of reaction, and it never reveals what the acceptance would have been if what you gave them had been better—if all the forces of art and creativity and daring and imagination had been unleashed. I believe in the people’s good sense and good taste, and I am not convinced that the people’s taste is as low as some of you assume....

Certainly I hope you will agree that ratings should have little influence where children are concerned. The best estimates indicate that during the hours of 5 to 6 p.m., 60 percent of your audience is composed of children under twelve. And most young children today, believe it or not, spend as much time watching television as they do in the schoolroom. I repeat—let that sink in—most young children today spend as much time watching television as they do in the schoolroom. It used to be said that there were three great influences on a child: home, school and church. Today there is a fourth great influence, and you ladies and gentlemen control it.

If parents, teachers, and ministers conducted their responsibilities by following the ratings, children would have a steady diet of ice cream, school holidays, and no Sunday school. What about your responsibilities? Is there no room on television to teach, to inform, to uplift, to stretch, to enlarge the capacities of our children? Is there no room for programs deepening their understanding of children in other lands? Is there no room for a children's news show explaining something about the world to them at their level of understanding? Is there no room for reading the great literature of the past, teaching them the great traditions of freedom? There are some fine children's shows, but

they are drowned out in the massive doses of cartoons, violence, and more violence. Must these be your trademarks? Search your consciences and see if you cannot offer more to your young beneficiaries whose future you guide so many hours each and every day.

What about adult programming and ratings? You know, newspaper publishers take popularity ratings too. The answers are pretty clear; it is almost always the comics, followed by the advice-to-the-lovelorn columns. But, ladies and gentlemen, the news is still on the front page of all newspapers, the editorials are not replaced by more comics, the newspapers have not become one long collection of advice to the lovelorn. Yet newspapers do not need a license from the government to be in business—they do not use public property. But in television—where your responsibilities as public trustees are so plain—the moment that the ratings indicate that Westerns are popular, there are new imitations of Westerns on the air faster than the old coaxial cable could take us from Hollywood to New York....

Let me make clear that what I am talking about is balance. I believe that the public interest is made up of many interests. There are many people in this great country, and you must serve all of us. You will get no argument from me if you say that, given a choice between a Western and a

symphony, more people will watch the Western. I like Westerns and private eyes too—but a steady diet for the whole country is obviously not in the public interest. We all know that people would more often prefer to be entertained than stimulated or informed. But your obligations are not satisfied if you look only to popularity as a test of what to broadcast. You are not only in show business; you are free to communicate ideas as well as relaxation. You must provide a wider range of choices, more diversity, more alternatives. It is not enough to cater to the nation's whims—you must also serve the nation's needs....

Let me address myself now to my role, not as a viewer but as chairman of the FCC....I want to make clear some of the fundamental principles which guide me.

First, the people own the air. They own it as much in prime evening time as they do at 6 o'clock Sunday morning. For every hour that people give you, you owe them something. I intend to see that your debt is paid with service.

Second, I think it would be foolish and wasteful for us to continue any worn-out wrangle over the problems of payola, rigged quiz shows, and other mistakes of the past....

Third, I believe in the free enterprise system. I want to see broadcasting improved and I want you to do the job....

Fourth, I will do all I can to help educational television. There are still not enough educational stations, and major centers of the country still lack usable educational channels....

Fifth, I am unalterably opposed to governmental censorship. There will be no suppression of programming which does not meet with bureaucratic tastes. Censorship strikes at the taproot of our free society.

Sixth, I did not come to Washington to idly observe the squandering of the public's airwaves. The squandering of our airwaves is no less important than the lavish waste of any precious natural resource....

What you gentlemen broadcast through the people's air affects the people's taste, their knowledge, their opinions, their understanding of themselves and of their world. And their future. The power of instantaneous sight and sound is without precedent in mankind's history. This is an awesome power. It has limitless capabilities for good—and for evil. And it carries with it awesome responsibilities—responsibilities which you and I cannot escape....

Acknowledgments

This report is the culmination of the effort of many individuals who contributed their considerable knowledge, experience, and creativity to the NAEP 2002 reading assessment. The assessment was a collaborative effort among staff from the National Center for Education Statistics (NCES), the National Assessment Governing Board (NAGB), Educational Testing Service (ETS), Westat, and Pearson Educational Measurement. Most importantly, NAEP is grateful to the students and school staff who made the assessment possible.

The NAEP 2002 reading assessment was funded through NCES, in the Institute of Education Sciences of the U.S. Department of Education. The Associate Commissioners of Education Statistics, Peggy Carr and Val Plisko, and the NCES staff—Arnold Goldstein, Steven Gorman, Carol Johnson, Andrew Kolstad, Taslima Rahman and Marilyn Seastrom—worked closely and collegially with the authors to produce this report.

The NAEP project at ETS is directed by Stephen Lazer and John Mazzeo, with assistance from John Barone. Sampling and data collection activities were conducted by Westat under the direction of Rene Slobasky, Nancy Caldwell, Keith Rust, and Dianne Walsh. Printing, distribution, scoring, and processing activities were conducted by Pearson Educational Measurement under the direction of Brad Thayer, Connie Smith, and William Buckles.

Test development activities took place at ETS under the direction of Patricia Donahue with assistance from Robert Finnegan.

The complex statistical and psychometric activities necessary to report results for the NAEP 2002 reading assessment were directed by Catherine Hombo and Jinming Zhang, with assistance from Andreas Oranje, Hui Deng, Kelvin Gregory, and Ying Jin.

The extensive data processing and computer programming activities underlying the statistical and psychometric analyses conducted at ETS are under the direction of David Freund, Edward Kulick, Bruce Kaplan, and Steven Isham. Data analyses presented in this report were managed by Steven Isham and Tatyana Petrovicheva with assistance from Laura Jerry, Youn-hee Lim, Norma Norris, Alfred Rogers, Fred Schaefer, Mike Weiss, and John Willey. The complex database work for this assessment was managed by Katherine Pashley with assistance from Gerry Kokolis.

The design and production of this report were overseen by Loretta Casalaina with assistance from Joseph Kolodey and Rick Hasney. Wendy Grigg coordinated the documentation and data checking procedures with assistance from Janice Goodis, Ming Kuang, Andrea Bergen, and Alice Kass. Arlene Weiner coordinated the editorial and proofreading procedures with assistance from Trish Hamill. The Web version of this report was coordinated by Rick Hasney with assistance from Loretta Casalaina. The consistency review process was coordinated by Pat O'Reilly.

Many thanks are due to the numerous reviewers, both internal and external to NCES and ETS. The comments and critical feedback of the following reviewers are reflected in the final version of this report: James Carlson, Young Chun, Doug Cochrane, Lawrence Feinberg, Ray Fields, Janet Jones, Mariann Lemke, Jean Osborn, Alan Vanneman, and Aileen Waters.

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

Postage and Fees Paid
U.S. Department of Education
Permit No. G-17

Official Business Only
Penalty for Private Use, \$300

