

A Comparison of the PIRLS 2011 and NAEP 2011 Fourth-Grade Reading Assessments

The purpose of this document is to provide background information that will be useful in interpreting the 2011 results from the Progress in Reading Literacy Study (PIRLS) by comparing its design, features, framework, and items with those of the U.S. National Assessment of Educational Progress.

The Progress in International Reading Literacy Study (PIRLS) 2011 and the National Assessment of Educational Progress (NAEP) 2011 reading assessment both measure the level of fourth-graders' reading literacy by asking students to read passages and answer accompanying reading comprehension questions. However, the two assessments were created for different purposes. PIRLS is meant to provide internationally comparable data on student reading abilities, while NAEP is meant to provide national, national subpopulation, and state and participating district level data on students' reading abilities or performance (i.e., what students know and can do). Thus, it is not surprising that previous comparison studies of PIRLS and NAEP have found distinctive differences between the two assessments (Stephens and Coleman 2007; Binkley and Kelly 2003; Binkley and Williams 1996; Binkley and Rust 1994).¹

The purpose of this paper is to examine how the PIRLS 2011 international assessment relates to the NAEP 2011 national reading assessment. This paper seeks to answer two primary questions:

- How is reading defined by each assessment broadly, and in terms of content and cognitive dimensions?
- What are the similarities and differences in how the content and cognitive dimensions are operationalized between the PIRLS 2011 and NAEP 2011 reading assessments?

Answers to these questions will provide background information that will be useful in interpreting the 2011 results from PIRLS by comparing its design, features, and framework with that of NAEP.

Following this introduction, an overview of both PIRLS and NAEP is provided. The next section of this paper explains how PIRLS and NAEP define reading literacy. The paper then reports how the two assessments compare in terms of their definitions of reading, their types of reading passages, and their test questions (items). The value of this detailed information about the similarities and differences between these two assessments is to help policymakers, educators, and the general public better interpret any differences in U.S. students' performance on PIRLS versus on NAEP.

Overview of PIRLS and NAEP

PIRLS is an international reading assessment measuring trends in reading literacy of fourth-grade students in different countries. It is administered by the International Association for the Evaluation of Educational Achievement (IEA) every 5 years. PIRLS 2011 is the third cycle of the PIRLS series, which began in 2001. The PIRLS 2011 Assessment Framework is grounded on the framework for PIRLS 2001; the aspects of reading being assessed have remained relatively similar throughout the 2001, 2006, and 2011 assessments. For PIRLS 2011, a total of 53 education systems participated, and the U.S. student sample included 12,726 fourth-graders from 370 schools who participated in the national sample and an additional 2,598 students from 77 schools who participated in a Florida benchmarking sample.

¹ The last two references cited reports on studies that compare NAEP with the predecessor of PIRLS, the IEA Reading Literacy Study.

The NAEP reading assessment measures reading comprehension skills of students in fourth grade. It is administered every other year to a nationally representative sample of U.S. students. The NAEP 2011 Reading Framework was revised most recently in 2009 and is thus different from the 1992–2007 frameworks. These differences include increasing emphasis on literary and informational texts, redefinitions of the cognitive processes involved in reading, addition of a new vocabulary assessment, and inclusion of poetry for fourth-grade students (National Assessment Governing Board [NAGB] 2010). The NAEP 2011 Reading Framework is the same as the NAEP 2009 Reading Framework. The NAEP 2011 reading assessment was administered to 211,000 fourth-grade students.² NAEP samples many more U.S. students than does PIRLS. In addition to being administered nationally, NAEP is also administered at the state level and within selected urban districts.

Comparison Between PIRLS 2011 and NAEP 2011 Reading Assessments

The focus of this paper is limited to a general comparison of the PIRLS 2011 and the NAEP 2011 fourth-grade reading assessments. This study adopted the same model used in the first PIRLS-NAEP comparison study by Binkley and Kelly (2003) and systematically examined the frameworks, reading passages, and items of the two reading assessments in terms of their definitions of reading, characteristics of the selected reading passages, cognitive dimensions involved in the items, and the format of the items.

A panel of nationally recognized reading experts (Expert Panel) familiar with both NAEP and PIRLS was convened to conduct a systematic comparison of the PIRLS 2011 and the NAEP 2011 reading assessments. (See appendix A for a list of Expert Panel members.) The Expert Panel first assessed differences and similarities using existing information about the framework for each assessment. Then, to further examine the level of correspondence between the two assessments, the Expert Panel classified the PIRLS 2011 passages and items using the NAEP 2011 reading framework; that is, the PIRLS 2011 passages and items were mapped onto the NAEP 2011 framework.

The Expert Panel conducted its work during a 2-day meeting November 2-3, 2011, in Washington, D.C. Upon arrival, each member of the Expert Panel received a binder with resources for completing his or her review. (See appendix B for the table of contents of the participant binder.) In addition, prior to undertaking their review, all members of the Expert Panel received a general orientation to the review process given by NCEES program staff. This orientation covered the purpose; overview of PIRLS and NAEP frameworks; objectives and organization of the 2011 comparison study; review of materials and description of procedures; and example and practice items. The Expert Panel was then divided into two working groups, and the facilitator for each working group continued the orientation with more specific instructions for completing the process. (See appendix C for the facilitators' guide used for this in-group orientation.)

To complement the Expert Panel's work, the research team documented information about the two assessments and analyzed and compared the lengths and readability levels of each passage in PIRLS and NAEP. The results of the comparative study completed by the Expert Panel and the research team are presented in the sections that follow.

² Number was rounded to the nearest 100.

Definitions of Reading³

The *PIRLS 2011 Assessment Framework* (Mullis et al. 2009) defines reading literacy as follows:

- “[T]he ability to understand and use those written language forms required by society and/or valued by the individual. Young readers can construct meaning from a variety of texts. They read to learn, to participate in communities of readers in school and everyday life, and for enjoyment.” (Mullis et al. 2009, p. 11)
- PIRLS’s view of a good reader is expressed in this excerpt: “Readers are regarded as actively constructing meaning and as knowing effective reading strategies and how to reflect on reading. They have positive attitudes toward reading and read for recreation... They can enjoy and gain information from the many multi-modal forms in which text is presented in today’s society.” (Mullis et al. 2009, p. 12)

In addition to the previous NAEP reading frameworks, three additional sources were key to developing the definition of reading in the NAEP 2011 Reading Framework (NAGB 2010): the PIRLS framework; the Program for International Student Assessment (PISA); and *Reading for Understanding: Toward an R&D Program in Reading Comprehension*, developed by the RAND Reading Study Group. In the *Reading Framework for the 2011 NAEP* (NAGB 2010), reading is defined as “an active and complex process that involves understanding written text, developing and interpreting meaning, and using meaning as appropriate to type of text, purpose, and situation” (NAGB 2010, p. 2). Each component of this definition is further explained as follows:

1. **Understanding written text:** “Readers attend to ideas and content in a text by locating and recalling information and by making inferences needed for literal comprehension of the text. In doing so, readers draw on their fundamental skills for decoding printed words and accessing their vocabulary knowledge” (NAGB 2010, pp. 2-3).
2. **Developing and interpreting meaning:** “Readers integrate the sense they have made of the text with their knowledge of other texts and with their outside experience. They use increasingly complex inferencing skills to comprehend information implied by a text” (NAGB 2010, p. 3).
3. **Using meaning:** “Readers draw on the ideas and information they have acquired from text to meet a particular purpose or situational need. The use of text may be as straightforward as knowing the time when a train will leave a particular station, or it may involve more complex behaviors such as analyzing how an author developed a character’s motivation or evaluating the quality of evidence presented in an argument” (NAGB 2010, p. 3).

Based on these definitions of reading, one may conclude that both PIRLS and NAEP stress that reading is an active process in which readers employ complex strategies to construct meaning from the text. Both definitions of reading also focus on how readers acquire information from the text or use the meaning to serve different purposes.

³ This section is based primarily on a review of Chapter 1–Overview of IEA’s PIRLS Assessment in the *PIRLS 2011 Assessment Framework* (Mullis et al. 2009) and Chapter One–Overview in the *Reading Framework for the 2011 NAEP* (NAGB 2010).

Nevertheless, these definitions also suggest that some differences do exist in the definition of reading in PIRLS 2011 and NAEP 2011. PIRLS focuses attention on readers' enjoyment and positive attitude toward reading. This motivational aspect is not explicitly indicated in the definition of reading in NAEP. A review of the definition of reading in NAEP appears to place more emphasis on the cognitive dimension of reading. Its definition specifically describes in detail the sequential nature of the reading process, such as how reading evolves from decoding printed words, drawing simple inferences, using complex inferencing skills to evaluate and analyze the meaning of the text, and using the ideas and information learned to meet specific purposes or needs. PIRLS, by comparison, includes in its definition an additional explicit focus on how readers utilize the text to participate in their communities or in society.

Reading Passages⁴

PIRLS and NAEP both attempt to select authentic reading materials that are similar to what students typically encounter in their everyday experiences. In PIRLS 2011, a total of 10 passages were included in the assessment. In NAEP 2011, there were 12 passages in the fourth-grade reading assessment, among which were four passages shared with the NAEP 2011 eighth-grade reading assessment. The passages in PIRLS 2011 and NAEP 2011 were compared by text type, length, and difficulty level.

Passage Text Type

Passage text types are associated with readers' purposes for reading. The PIRLS assessment focuses on two reading purposes: reading for literary experience and reading to acquire and use information. PIRLS stresses that these two reading purposes do not align strictly with types of texts, so they use the general terms *literary texts* and *informational texts* to describe the text types of its passages. The proportion of literary and informational passages is set to be equal in PIRLS; the PIRLS 2011 assessment consisted of five literary passages and five informational passages.

NAEP also includes literary passages and informational passages as the two main text types. NAEP establishes clear distinctions in text types within literary texts and informational texts based on the structural characteristics of texts. Literary texts include three subtypes: fiction, literary nonfiction (essays, speeches, and autobiographies or biographies), and poetry. Informational texts also include three subtypes: exposition, argumentation and persuasive text, and procedural text and documents.

One additional point bears mentioning in this context. NAEP classifies autobiographies and biographies as a type of literary text because they follow a story structure and employ literary devices to present information and ideas. However, in PIRLS, autobiographies and biographies are considered a type of informational text.

As part of its comparative analysis of the two assessments, the Expert Panel classified the ten PIRLS 2011 (fourth-grade) passages using the NAEP 2011 reading framework. The results of the analyses were as follows: six of the PIRLS passages were classified as literary texts and four as informational texts in the NAEP 2011 framework. Of the six PIRLS passages classified as literary texts, five were further

⁴ This section is based primarily on a review of Chapter 2—PIRLS Reading Purposes and Processes of Reading Comprehension and Chapter 4—Assessment Design and Specifications in the *PIRLS 2011 Assessment Framework* (Mullis et al. 2009) and Chapter Two—Content and Design of NAEP in Reading in the *Reading Framework for the 2011 NAEP* (NAGB 2010). In addition, reported here are objective data and information about the PIRLS 2011 and NAEP 2011 fourth-grade reading assessments.

categorized as fiction and one as literary nonfiction (a biographical passage about Leonardo da Vinci). Of the four PIRLS passages classified as informational texts, one was categorized into the procedural text and documents subtype, and the remaining three were placed into the exposition subtype.

Table 1 presents the distribution of PIRLS 2011 passages and NAEP 2011 passages categorized by the NAEP 2011 framework. In NAEP 2011, 58 percent of the passages were literary texts (seven) and 42 percent were informational texts (five). When PIRLS passages were classified using the NAEP framework, there was a 60 percent-40 percent split between literary (six) and informational (four) passages.

Table 1. PIRLS 2011 and NAEP 2011 fourth-grade passage text types classified by NAEP 2011 Reading Framework

	Literary texts			Informational texts		
	Fiction	Literary nonfiction	Poetry	Exposition	Argumentation and persuasive	Procedural and documents
PIRLS	5	1	0	3	0	1
NAEP	4	1	2	5	0	0

SOURCE: U.S. Department of Education, National Center for Education Statistics, *A Comparison of the PIRLS 2011 and NAEP 2011 Fourth-Grade Reading Assessments*, 2012.

The Expert Panel noted that PIRLS 2011 does not have paired texts of different types in its passages but that NAEP 2011 does. While PIRLS involves only one text type for each passage, some NAEP passages are composed of two stories or two different types of text on a similar topic. For the NAEP fourth-grade assessment, the paired texts are usually two narrative stories or a poem with another piece of text. This passage format allows readers to compare and contrast ideas and information in the two texts and draw cross-text arguments.

Passage Length

Comparing the number of words in the PIRLS 2011 and NAEP 2011 passages revealed that the lengths of the PIRLS passages ranged from 570 words to 880 words, with an average of 802 words (table 2). In contrast, the passages in NAEP, excluding the two poems (47 words and 197 words) since PIRLS does not include poems, ranged from 473 words to 1,147 words, with an average word length of 840 words. Thus, excluding poetry, the passages in NAEP present students with more text to read than those in PIRLS. However, if poetry is included in the analysis, fourth-grade students taking the NAEP assessment who get poems might have less text to read (average word length of 721) than fourth-grade students taking the PIRLS assessment (average word length of 802). Table 2 also includes an identical scale between 0 and 1,147 for each assessment, where the data points are scaled across to represent the number of words in each passage. A comparison of the two scales shows a wider variation in the length of the NAEP passages.

Table 2. Number of words in the PIRLS 2011 and NAEP 2011 fourth-grade reading passages

	Number of words in each text	Average	Average without poetry
PIRLS texts	570, 767, 768, 770, 811, 855, 860, 869, 870, 880	802	802
	0 ————— 1,147		
NAEP texts	47, 197, 473, 782, 796, 808, 809, 878, 885, 908, 918, 1,147	721	840
	0 ————— 1,147		

SOURCE: U.S. Department of Education, National Center for Education Statistics, *A Comparison of the PIRLS 2011 and NAEP 2011 Fourth-Grade Reading Assessments*, 2012.

Passage Difficulty

To compare the difficulty levels of the PIRLS 2011 and NAEP 2011 passages, a number of readability measures were used to assess each text in the two assessments. These included the Flesch Reading Ease Formula, the Flesch-Kincaid Grade Level, the Lexile scores, and the Fry Graph. These measures were selected from those for which data were provided by NAEP and in order to align with those also used by the TIMSS & PIRLS International Study Center. In general, these approaches share similar assumptions on measuring readability. They use sentence lengths, total number of words, and word lengths to estimate the syntactic complexity and semantic difficulty of the text. These measures also match text difficulty with readers' age or grade level and provide an estimate of the appropriate readability age or grade level for a text. A caution to note is that the readability formulas are used as a quick assessment of the difficulty of a text and do not account for certain features of the text that also have influences on reading comprehension, such as text structure, topic, and appeal.

Tables 3 through 6 present the approximate grade levels or readability scores of each passage in PIRLS 2011 and the NAEP 2011 reading assessment determined using these four readability measures. These tables also include identically scaled and sized lines to give a visual image of the dispersion of the grade levels or readability scores for the two assessments. The overall average grade level or average readability score for each assessment was derived by taking the average of the grade level or readability scores for all of the passages within each assessment. One point to note is that the poems in NAEP were excluded from these analyses. Poetry is a unique genre different from prose; its readability does not depend solely on sentence lengths, word lengths, or word frequencies, and the readability approaches used here are not an adequate assessment of level of difficulty for poetry. Therefore, to ensure an unbiased comparison between PIRLS and NAEP, the two poems in NAEP were not included in the calculation of the average readability scores.

The **Flesch Reading Ease Formula** takes into consideration the total number of words, the number of syllables, and the total number of sentences. Unlike most formulas, which report their results as grade-level scores, the Flesch Reading Ease Formula reports its results on a difficulty scale ranging from 0 to 100; the higher the score, the easier the material. The readability scores correspond to the following readability levels: very difficult (0-29), difficult (30-49), fairly difficult (50-59), standard (60-69), fairly easy (70-79), easy (80-89), and very easy (90-100). As shown in table 3, the average Flesch Reading Ease Formula score for PIRLS 2011 is 83.6 (easy) and for NAEP 2011 is 76.5 (fairly easy), indicating that by comparison the PIRLS passages are generally easier than the NAEP passages.

Table 3. Readability scores determined by the Flesch Reading Ease Formula for the PIRLS 2011 and NAEP 2011 fourth-grade reading passages

	The Flesch Reading Ease Score	Average score
PIRLS	66, 78, 81, 82, 83, 83, 89, 90, 92, 92	83.6
NAEP	62, 64, 68, 74, 75, 77, 79, 80, 89, 97	76.5

SOURCE: U.S. Department of Education, National Center for Education Statistics, *A Comparison of the PIRLS 2011 and NAEP 2011 Fourth-Grade Reading Assessments*, 2012.

Similar to the Flesch Reading Ease measure, the **Flesch-Kincaid Grade Level** measure is based on the number of words, syllables, and sentences in a text, but with slightly different weighting. This measure matches the text with an approximate readability grade level. A higher grade level indicates a more difficult text. As shown in table 4, the Flesch-Kincaid Grade Level measure predicts that the grade range for the PIRLS 2011 passages is between grade 2 and grade 7, with an average grade level of 4.5. Additionally, the grade range for NAEP 2011 passages is between grade 1 and grade 8, with an average grade level of 5.9. Based on the Flesch-Kincaid Grade Level measure, on average, the PIRLS 2011 passages would be more appropriate for students between fourth grade and fifth grade. By comparison the NAEP 2011 passages would be considered appropriate for students at the end of fifth grade.

Table 4. Readability grade levels determined by the Flesch-Kincaid Grade Level measure for the PIRLS 2011 and NAEP 2011 fourth-grade reading passages

	The Flesch-Kincaid grade level	Average grade level
PIRLS	2.5, 2.8, 3.2, 3.5, 4.5, 4.9, 5.1, 5.3, 5.5, 7.6	4.5
NAEP	1.3, 4.5, 4.8, 5.9, 5.9, 6.4, 7, 7.1, 7.4, 8.4	5.9

SOURCE: U.S. Department of Education, National Center for Education Statistics, *A Comparison of the PIRLS 2011 and NAEP 2011 Fourth-Grade Reading Assessments*, 2012.

The **Lexile** analysis, shown in table 5, takes into account the sentence length and word frequency of a text to determine its readability level. Higher Lexile scores represent a higher level of complexity of the text. In general, the results suggest that the comprehension of the PIRLS 2011 passages requires reading abilities at a lower grade level than required for the NAEP 2011 passages. The corresponding average grade level for the PIRLS 2011 passages is between fourth and fifth grade, whereas the corresponding average grade level for the NAEP 2011 passages is at sixth grade.

Table 5. Readability scores from the Lexile analysis for the PIRLS 2011 and NAEP 2011 fourth-grade reading passages

	Lexile analysis (Lexile scores)	Average Lexile score	Corresponding grade level ¹
PIRLS	480, 560, 640, 690, 800, 830, 910, 920, 1010, 1,030	787	Fourth to fifth grade
NAEP	360, 800, 890, 940, 950, 990, 1,000, 1010, 1,080, 1,080	910	Sixth grade

¹ There is no direct correspondence between a Lexile score and a specific grade level. The text complexity in the range from the 25th percentile to the 75th percentile of typical texts designed for fourth-graders is between 640L and 780L, for fifth-graders between 730L and 850L, and for sixth-graders between 860L and 920L.

SOURCE: U.S. Department of Education, National Center for Education Statistics, *A Comparison of the PIRLS 2011 and NAEP 2011 Fourth-Grade Reading Assessments*, 2012.

The readability grade level from the **Fry Graph** measure, as shown in table 6, is based on the average number of sentences and average number of syllables in a text. Ranging from level 1 to level 15, the Fry Graph measure predicts the grade level required to comprehend a text. Table 6 shows that the results from the Fry Graph analysis were similar to those from the previous readability analyses. The grade range for the PIRLS 2011 passages is between third grade and eighth grade, and the grade range for the NAEP 2011 passages is between first grade and tenth grade. The average Fry Graph grade level for the PIRLS 2011 passages is nearly fifth grade, whereas the average Fry Graph grade level for the NAEP 2011 reading passages is nearly seventh grade.

Table 6. Readability grade levels from the Fry Graph analysis for the PIRLS 2011 and NAEP 2011 fourth-grade reading passages

	Fry Graph (Fry grade level)	Average grade level
PIRLS	3, 3, 3, 3, 5, 5, 5, 6, 6, 8	4.7
NAEP	1, 5, 5, 7, 7, 7, 8, 9, 10, 10	6.9

SOURCE: U.S. Department of Education, National Center for Education Statistics, *A Comparison of the PIRLS 2011 and NAEP 2011 Fourth-Grade Reading Assessments*, 2012.

These readability analyses indicate that the difficulty levels in PIRLS 2011 and NAEP 2011 reading passages vary by degrees; the PIRLS passages have a narrower range of variation in terms of reading difficulty than do NAEP passages. Taken together, these readability analyses suggest that the PIRLS 2011 passages are, on average, shorter and about one grade level lower than the NAEP 2011 passages.

Items⁵

Both PIRLS and NAEP assess students' reading skills by the design of comprehension questions (also known as items). In each assessment, students are presented with a set of questions to which they must respond after they finish reading a passage. The items in the PIRLS 2011 and NAEP 2011 reading assessments were compared by the Expert Panel with respect to the cognitive processes being assessed and item formats.

Cognitive Processes

In PIRLS, four comprehension processes are assessed:

1. **Focusing on and retrieving explicitly stated information:** “Successful retrieval requires a fairly immediate or automatic understanding of the text. This process needs little or no inferring or interpreting. There are no ‘gaps’ in meaning to be filled—the meaning is evident and stated in the text.... Focus on the text typically remains at the sentence or phrase level in this type of text processing” (Mullis et al. 2009, p. 25).
2. **Making straightforward inferences:** “Making inferences allows the reader to move beyond the surface of texts and to fill in the ‘gaps’ in meaning that often occur in texts. Some of these inferences are straightforward in that they are based mostly on information that is contained in the text: the reader may merely need to connect two or more ideas or pieces of information” (Mullis et al. 2009, pp. 25-26).
3. **Interpreting and integrating ideas and information:** “[T]he reader is processing text beyond the phrase or sentence level.... When they interpret and integrate text information and ideas, readers may need to draw on their background knowledge and experiences more than they do for straightforward inferences. Because of this, meaning that is constructed through interpreting and integrating ideas and information is likely to vary among readers, depending upon the experiences and knowledge they bring to the reading task” (Mullis et al. 2009, p. 27).
4. **Examining and evaluating content, language, and textual elements:** “[T]he focus shifts from constructing meaning to critically considering the text itself.... The reader engaged in this process is standing apart from the text and examining or evaluating it. The text content, or meaning, may be examined from a very personal perspective or with a critical and objective view. Here the reader relies on knowledge about the world or on past reading.... [R]eaders draw upon their knowledge of language usage and general or genre-specific features of texts” (Mullis et al. 2009, p. 28).

The PIRLS 2011 assessment includes a total of 142 items. The distribution of the items across the four comprehension processes in PIRLS 2011 is as follows: 23 percent in “Focusing on and retrieving explicitly stated information,” 32 percent in “Making straightforward inferences,” 32 percent in

⁵ This section is based primarily on a review of Chapter 2—PIRLS Reading Purposes and Processes of Reading Comprehension and Chapter 4—Assessment Design and Specifications in the *PIRLS 2011 Assessment Framework* (Mullis et al. 2009) and Chapter Two—Content and Design of NAEP in Reading in the *Reading Framework for the 2011 NAEP* (NAGB 2010). In addition, reported here are objective data and information about the PIRLS 2011 and NAEP 2011 fourth-grade reading assessments.

“Interpreting and integrating ideas and information,” and 13 percent in “Examining and evaluating content, language, and textual elements.”

In NAEP, the term “cognitive target” refers to the mental processes underlying reading comprehension. NAEP assesses three cognitive targets:

1. **Locating and recalling:** “As students locate or recall information from what they read, they may identify clearly stated main ideas or supporting details or they may find essential elements of a story, such as characters, time, or setting. Their process in answering assessment items often involves matching information given in the item to either literal or synonymous information in the text before they can then use the textual information to develop a response” (NAGB 2010, p. 37).
2. **Integrating and interpreting:** “When readers engage in behaviors involving integrating and interpreting, they make comparisons and contrasts of information or character actions, examine relations across aspects of text, or consider alternatives to what is presented in text. This aspect of the reading is critical to comprehension and can be considered the stage in which readers really move beyond the discrete information, ideas, details, themes, and so forth presented in text and extend their initial impressions by processing information logically and completely” (NAGB 2010, p. 37).
3. **Critiquing and evaluating:** “The focus remains on the text itself but the reader’s purpose is to consider the text critically by assessing it from numerous perspectives and synthesizing what is read with other texts and other experiences. Items may ask students to evaluate the quality of the text as a whole, to determine what is most significant in a passage, or to judge the effectiveness of specific textual features to accomplish the purpose of the text.... Items might ask for the likelihood that an event could actually have taken place, the plausibility of an argument, or the adequacy of an explanation for an event” (NAGB 2010, p. 38).

The Expert Panel mapped the 142 items from the PIRLS 2011 assessment to the three cognitive targets in NAEP 2011 framework, reaching consensus on the final coding. The results are presented in table 7, which shows the distribution of PIRLS 2011 and NAEP 2011 items across NAEP’s cognitive targets. The first row presents the results of the mapping of PIRLS 2011 items to the NAEP cognitive targets by the Expert Panel. The second row shows the actual distribution of NAEP 2011 items by cognitive target according to the NAEP assessment developers’ classification.

Table 7. Distribution of the PIRLS 2011 and NAEP 2011 fourth-grade items by NAEP cognitive targets

	Locate and recall	Integrate and interpret	Critique and evaluate
PIRLS items (n=142)	50% (71)	46% (66)	4% (5)
NAEP items (n=120)	17% (20)	69% (83) ¹	14% (17)

¹ This category includes 20 meaning vocabulary items. The vocabulary items are multiple-choice questions that ask readers to select the most appropriate meaning for a word as it is used in the context of a passage from four possible options. PIRLS does not have multiple-choice vocabulary items.

NOTE: Details might not sum to 100 percent because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, *A Comparison of the PIRLS 2011 and NAEP 2011 Fourth-Grade Reading Assessments*, 2012.

All of the PIRLS 2011 items could be mapped to the three cognitive targets of NAEP. A total of 50 percent of the PIRLS 2011 items were classified as “locate and recall,” 46 percent as “integrate and interpret,” and 4 percent as “critique and evaluate” in the NAEP 2011 framework. Compared with NAEP 2011, PIRLS 2011 has more “locate and recall” items and fewer “critique and evaluate” items. This reflects the fact that the NAEP assessment focuses on how readers integrate, interpret, or evaluate the text, while PIRLS focuses more on readers’ text-based understanding. It should be noted that the NAEP “integrate and interpret” category includes vocabulary items. According to the *Reading Framework for the 2011 NAEP* (NAGB 2010), vocabulary assessment occurs in the context of a passage, and functions as a measure of passage comprehension and as a test of readers’ knowledge of the word’s meaning.

Tables 8 through 11 provide detailed information about how the four comprehension processes in PIRLS were mapped to the three cognitive targets in NAEP. This permits a more complete examination of the correspondence between the PIRLS comprehension processes and the NAEP cognitive targets.

Focusing on and retrieving explicitly stated information. Table 8 shows that all of the PIRLS 2011 items designed to measure the comprehension process “focus on and retrieve explicitly stated information” fit the NAEP 2011 cognitive target “locate and recall.” This suggests that both of these two processes are designed to measure the same underlying cognitive dimension.

Table 8. PIRLS 2011 “Focus on and Retrieve Explicitly Stated Information and Ideas” items classified by NAEP cognitive targets

NAEP cognitive target	Panel classification of PIRLS “Focus on and Retrieve Explicitly Stated Information and Ideas” items
Total	100%
Locate and recall	100%
Integrate and interpret	0%
Critique and evaluate	0%

SOURCE: U.S. Department of Education, National Center for Education Statistics, *A Comparison of the PIRLS 2011 and NAEP 2011 Fourth-Grade Reading Assessments*, 2012.

Making straightforward inferences. Table 9 indicates that PIRLS 2011 items designed to measure the comprehension process of “making straightforward inferences” overlap with both the NAEP 2011 cognitive target “locate and recall” (57 percent) and “integrate and interpret” (43 percent). The Expert Panel mentioned that the definition of NAEP’s “locate and recall” not only involves identifying and locating textually explicit information in the text but also includes making simple and straightforward inferences. Therefore, if the items in the PIRLS comprehension process of “make straightforward inferences” require readers to carry out simple inferences from the information explicitly presented in the passage, they tended to classify these as NAEP’s “locate and recall.”

Table 9. PIRLS 2011 “Make Straightforward Inferences” items classified by NAEP cognitive targets

NAEP cognitive target	Panel classification of PIRLS “Making Straightforward Inferences” items
Total	100%
Locate and recall	57%
Integrate and interpret	43%
Critique and evaluate	0%

SOURCE: U.S. Department of Education, National Center for Education Statistics, *A Comparison of the PIRLS 2011 and NAEP 2011 Fourth-Grade Reading Assessments*, 2012.

Interpreting and integrating ideas and information. Results in table 10 reveal that of the PIRLS 2011 items measuring the comprehension process of “interpret and integrate ideas and information,” 78 percent were classified as “integrate and interpret,” 20 percent as “locate and recall,” and 2 percent as “critique and evaluate” using NAEP’s 2011 cognitive targets. These results suggest that, overall, items classified in PIRLS as “interpret and integrate ideas and information” overlap to a considerable extent with the NAEP cognitive target “integrate and interpret.” However, 20 percent of the PIRLS “interpret and integrate” items were classified by the Expert Panel at the more basic NAEP cognitive target of “locate and recall.”

Table 10. PIRLS 2011 “Interpret and Integrate Ideas and Information” items classified by NAEP cognitive targets

NAEP cognitive target	Panel classification of PIRLS “Interpret and Integrate Ideas and Information” items
Total	100%
Locate and recall	20%
Integrate and interpret	78%
Critique and evaluate	2%

SOURCE: U.S. Department of Education, National Center for Education Statistics, *A Comparison of the PIRLS 2011 and NAEP 2011 Fourth-Grade Reading Assessments*, 2012.

Examining and evaluating content, language, and textual elements. Table 11 shows that of the PIRLS 2011 items measuring the comprehension process “examine and evaluate content, language, and textual elements,” 61 percent were classified by the Expert Panel as “integrate and interpret,” 22 percent as “critique and evaluate,” and 17 percent as “locate and recall” in NAEP 2011. These results suggest that PIRLS “examine and evaluate content, language, and textual elements” items align better with NAEP’s “integrate and interpret” than with NAEP’s “critique and evaluate” cognitive target.

Table 11. PIRLS 2011 “Examine and Evaluate Content, Language, and Textual Elements” items classified by NAEP cognitive targets

NAEP cognitive target	Panel classification of PIRLS “Examine and Evaluate Content, Language, and Textual Elements” items
Total	100%
Locate and recall	17%
Integrate and interpret	61%
Critique and evaluate	22%

SOURCE: U.S. Department of Education, National Center for Education Statistics, *A Comparison of the PIRLS 2011 and NAEP 2011 Fourth-Grade Reading Assessments*, 2012.

Although “examine and evaluate content, language, and textual elements” is the highest cognitive level in PIRLS 2011, it still taps more of the skills assessed by the second cognitive level of NAEP 2011 (“integrate and interpret”) rather than the highest level in NAEP 2011 (“critique and evaluate”). Moreover, 17 percent of the PIRLS “examine and evaluate content, language, and textual elements” items were classified at NAEP’s basic level “locate and recall.” Most of these items are constructed-response questions that ask readers to find explicit information in the text and copy down the exact words on the answer sheet. The Expert Panel considered this type of item as only measuring readers’ skills in locating literal information from the text.

The finding that PIRLS “examine and evaluate content, language, and textual elements” distributes widely across the NAEP 2011 cognitive targets implies a key difference in how PIRLS 2011 and NAEP 2011 assess the examining and evaluating cognitive process of reading. After reviewing all the PIRLS

items, the Expert Panel pointed out that the items defined by PIRLS as examining and evaluating, collectively, do not resemble the ways in which NAEP assesses evaluation skills in reading. The panel explained that typical items in NAEP’s cognitive target of “critique and evaluate” would ask readers to select from several plausible alternatives and provide justification to defend their choice, or require readers to offer a critique and explain the criteria for their arguments. PIRLS does not contain these types of items; generally, readers are able to draw responses directly from the passages. The panel also felt that PIRLS items focus more on examining the text, rather than evaluating the text. Most of the PIRLS items ask readers to examine existing information in the text in more detail. PIRLS readers are not required to generate criteria for evaluation or to take a stand and provide convincing support. Therefore, the Expert Panel concluded that the design of the PIRLS 2011 items is, in general, less cognitively demanding than are NAEP 2011 items.

In summary, the Expert Panel’s classification and discussion of the PIRLS 2011 items by cognitive dimension reveal striking differences between PIRLS 2011 and NAEP 2011. PIRLS has more “locate and recall” items than those in NAEP, while NAEP has more “integrate and interpret” items and also more “critique and evaluate” items than exist in the PIRLS assessment. In addition, NAEP’s “critique and evaluate” items request a more complex cognitive process than is required for most of the PIRLS items that are also classified as “critique and evaluate.” NAEP items in this category ask readers to offer persuasive personal opinions by incorporating information outside the text, such as their background knowledge or individual perspectives, and relating that information to the text. By contrast, PIRLS items classified as “critique and evaluate” tend to assess text-based understanding and focus on how readers retrieve and examine information within the text.

Item Format

In PIRLS 2011 and NAEP 2011, passages are accompanied by a set of questions (items) in multiple-choice and constructed-response formats. In both assessments, multiple-choice items provide four possible response options, of which only one is correct. Constructed-response questions include short-answer items and extended-response items. Short-answer items are worth one or two points and extended-response items are awarded up to three points, depending on the level of accuracy and thoroughness of the written responses. To establish a reliable scoring procedure for scorers, each constructed-response item has a scoring guide that describes the essential characteristics of appropriate and complete responses for each score level.

In PIRLS 2011, the proportion of multiple-choice items and constructed response items are nearly equal; 52 percent of the items are in multiple-choice format and 48 percent of the items are in constructed-response format. The NAEP 2011 reading assessment, on the other hand, has many more multiple-choice items (70 percent) than constructed-response items (30 percent). One of the reasons for the higher percentage of multiple-choice items in NAEP 2011 is that the NAEP reading assessment includes a block of items to measure students’ vocabulary, and these items are all in multiple-choice format. The distribution of item types in PIRLS 2011 and NAEP 2011 are presented in table 12.

Table 12. Distribution of item formats in PIRLS 2011 and NAEP 2011 fourth-grade reading assessments

	Multiple choice	Constructed response
PIRLS	52%	48%
NAEP	70%	30%

SOURCE: U.S. Department of Education, National Center for Education Statistics, *A Comparison of the PIRLS 2011 and NAEP 2011 Fourth-Grade Reading Assessments*, 2012.

The Expert Panel also identified some qualitative differences in the requirement and structure of the constructed-response items between PIRLS 2011 and NAEP 2011. One major distinction was that PIRLS has constructed-response items for which full points could be earned by copying exact words or sentences from the text, while NAEP does not. The Expert Panel explained that constructed-response items in NAEP are used to assess how readers interpret, integrate, evaluate, or critique the text but not to assess whether readers can identify relevant information in the text.

The Expert Panel also noted that PIRLS constructed-response items offer a scaffolding structure for test-takers. In PIRLS 2011, if the item asks for more than one piece of information about the text, it numbers each response needed separately (for example, 1: _____, 2: _____) on the answer sheet in advance, so readers just need to fill in after each number when responding. NAEP 2011 items do not have this feature. Moreover, the panel noted that PIRLS 2011 sometimes uses pictures or graphics as prompts or signals to direct readers to a specific part of the text being asked about, while NAEP 2011 does not use any pictorial prompts or signals in its items. In effect, NAEP does not provide any cues in its items to help readers locate a particular section of text. These differences tend to make NAEP 2011 more cognitively demanding than PIRLS 2011.

Summary

In the United States, data on fourth-grade students' reading achievement come primarily from two sources: the National Assessment of Educational Progress (NAEP) and the Progress in International Reading Literacy Study (PIRLS). The purpose of this current study was to examine how the PIRLS 2011 international assessment relates to the NAEP 2011 national reading assessment. The study examined how reading is defined by each assessment broadly, and in terms of content and cognitive dimensions. Also, NCES compared the form and content of the PIRLS and NAEP reading assessments. The results of this study provide background information that is useful in interpreting the 2011 results from PIRLS by comparing its design, features, and framework with that of NAEP.

This comparative study of PIRLS 2011 and NAEP 2011 revealed important similarities and differences between the two assessments. Both the PIRLS and NAEP assessments have a similar definition of reading. Both define reading literacy as an active and constructive process between readers and texts, and both emphasize how readers draw connections across sentences and interpret meanings in the text. The two assessments also employ literary texts and informational texts as the main text types of the reading passages used in the assessments. In addition, both assessments involve two types of items: multiple choice and constructed response. However, there are important differences between the PIRLS 2011 and NAEP 2011 reading assessments, which are summarized below.

Passages text type analyses reveal that PIRLS 2011 and NAEP 2011 have relatively equal proportions of literary texts and informational texts and that both assessments have more literary texts than informational texts. However, NAEP 2011 included poetry in its assessment, whereas PIRLS 2011 did not. NAEP 2011 also included paired texts in its assessment in which readers compare two different texts on a similar topic simultaneously, while PIRLS 2011 readers only read one text at a time.

In examining passage length and difficulty, PIRLS 2011 passages were shorter on average than the NAEP 2011 passages. Readability analyses indicate that, on average, the PIRLS 2011 passages were about one grade level lower than the NAEP 2011 passages. However, it should be noted that of the twelve passages available for the NAEP 2011 fourth-grade reading assessment three were shared with the NAEP 2011 eighth-grade reading assessment.

Item-by-item content review also showed some differences between the assessments. About half of the PIRLS 2011 items were mapped to the NAEP "locate and recall" cognitive target. Most of the remaining PIRLS 2011 items were mapped to the NAEP "integrate and interpret" cognitive target. Very few items were mapped to the NAEP "critique and evaluate" cognitive target. By contrast, NAEP 2011 had more items to assess the "integrate and interpret" as well as the "critique and evaluate" cognitive targets than did PIRLS 2011. The comparison on the cognitive dimensions measured in each assessment indicates that PIRLS 2011 focused more on assessing readers' skills in analyzing information within the text and drawing text-based inferences, whereas NAEP 2011 placed more emphasis on how readers develop inferences and personal interpretations by utilizing personal knowledge or perspectives to examine and evaluate the text in relation to that knowledge or perspectives.

Although both assessments used both multiple-choice and constructed-response items, the constructed-response items in PIRLS 2011 listed separately on the answer sheet the number for each written response needed as a way to scaffold the answering process for readers. PIRLS 2011 also used pictures or symbols within the text to cue test-takers to a specific part of the text where information for answers could be found. These features were absent in NAEP; NAEP 2011 did not provide a scaffolding structure, nor did it offer cues in the form of visual aids to help test-takers.

In summary, there are distinctive differences between PIRLS 2011 and NAEP 2011. Overall, these differences suggest that the NAEP 2011 reading assessment may be more cognitively challenging than PIRLS 2011 for U.S. fourth-grade students. Taken together, these findings suggest that caution should be exercised when attempting to compare fourth-grade students' performance on PIRLS 2011 with fourth-grade students' performance on the NAEP 2011 reading assessment.

References

- Binkley, M., and Kelly, D. L. (2003). *A Content Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments* (NCES 2003-10). U.S. Department of Education. Washington, DC: National Center for Education Statistics (NCES).
- Binkley, M., and Rust, K. (Eds.). (1994). *Reading Literacy in the United States: Technical Report of the U.S. Component of the IEA Reading Literacy Study* (NCES 94-259). U.S. Department of Education. Washington, DC: National Center for Education Statistics (NCES).
- Binkley, M., and Williams, T. (1996). *Reading Literacy in the United States: Findings from the IEA Reading Literacy Study* (NCES 96-258). U.S. Department of Education. Washington, DC: National Center for Education Statistics (NCES).
- Mullis, I. V.S., Martin, M. O., Kennedy, A. M., Trong, K. L., and Sainsbury, M. (2009). *PIRLS 2011 Assessment Framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- National Assessment Governing Board (NAGB). (2010). *Reading Framework for the 2011 National Assessment of Educational Progress*. Washington, DC: U.S. Government Printing Office.
- Stephens, M., and Coleman, M. (2007). *Comparing PIRLS and PISA with NAEP in Reading, Mathematics, and Science*. U.S. Department of Education. Washington, DC: National Center for Education Statistics Working Paper. Retrieved June 20, 2012, from <http://nces.ed.gov/Surveys/PISA/pdf/comppaper12082004.pdf>.

Appendix A

Comparison Between PIRLS 2011 and NAEP 2011 Reading Expert Panel Review Meeting Expert Panel Members

Patricia Alexander
University of Maryland, College Park

Mary Beth Curtis
Lesley University

Margaret McKeown
University of Pittsburgh

Bonnie Meyer
Pennsylvania State University

Appendix B

Comparison Between PIRLS 2011 and NAEP 2011 Reading Expert Panel Review Meeting Participant Binder Table of Contents

- Tab 1** Agenda for the Reading Meeting
- Tab 2** Reading Expert Panel Members: Participant List
- Tab 3** Training Presentation Slides:
Comparison Between PIRLS 2011 and NAEP 2011: The Reading
Assessment for Grade 4
- Tab 4** Results from Previous Comparison Study:
A Content Comparison of the NAEP 2002 and PIRLS 2001 Fourth-
Grade Reading Assessment
- Tab 5** 2011 NAEP Reading Framework Summary
- Tab 6** Comparison Between PIRLS 2011 and NAEP 2011:
Discussion Guide for the Reading Assessment for Grade 4
- Tab 7** Item Classification Forms — Literary Texts
- Tab 8** Item Classification Forms — Informational Texts

Appendix C

Comparison Between PIRLS 2011 and NAEP 2011 Facilitators' Guide for Orientation of Expert Panel

Primary Duties

1. Moderate the item classification process.
 - At the beginning, lead the group to classify the items one-by-one for the first passage.
 - Thereafter, let the panel classify the second passage into text types.
 - Then, they will work on the whole item set for that passage individually.
 - After each panel member finishes their individual work, have them check the results as a group and discuss to reach group consensus.

Introduction

2. To begin with, summarize the objective of this expert meeting:

“From the presentation, we know that the purpose of this study is to identify the differences and similarities between the frameworks of NAEP and PIRLS in terms of text types and cognitive processes of these two assessments. Within the small group, we are going to map the PIRLS items into the NAEP framework. You can see that in your binder, you have the framework summaries with codes for text types and cognitive targets. You also have item review sheets in the binder and the item set. Please take out the item review sheets and open the item binder to the first item of this passage. We are going to classify the first set of items one by one to get familiar with this process. Please match the first item into the NAEP framework in terms of the cognitive target on the first row of the item review sheet. You should also decide whether the item is a grade-equivalent task for a fourth-grader in U.S. curricula.”

Review of Resources/Documents Available

3. Review the resources available to members of the panel.
 - Participant Binders
 - TAB 1: Agenda
 - TAB 2: Participants List
 - TAB 3: Copy of Training Presentation (Slides)
 - TAB 4: Report from Previous Comparison Study
 - TAB 5: Synopses of 2011 NAEP Framework Content Specifications
 - TAB 6: Discussion Guide
 - TABS 7-8: Item Classification Forms by text type

- In-Room Binders
 - Items divided by text type.
 - Within each type, items are ordered by passage.
 - The item set is placed after the passages, and the scoring rubrics are attached after the item set. The item review sheets have a column that tells you whether the rubrics for certain items are available.
 - Each participant will have one PIRLS framework binder and one NAEP framework binder.
- PIRLS Item Information Spreadsheets

Classification of the 1st Item

4. “Since we have done the passage classification for the first passage (Flowers on the Roof) and the first item in the practice session, now let’s try the second item.”
 - Give the panel 5 minutes to read the whole passage again.
 - Give the panel 2-3 minutes to finish the second item **individually**.
 - Then, **as a group**, ask them to check the answers with each other by asking: “What is the cognitive target of Item 2?”
 - **Confirm the group consensus for each item at the end by asking:**
 - “Do we all agree that this item:
 - ✓ Matches the cognitive target _____, and
 - ✓ _____ (<<Does or Does not>>) fit within U.S. curricula?”

Classification of the REMAINING Items for the First Passage

5. Repeat the same process for the rest of the items, **one at a time**.
 - **Prior to the Expert Panel meeting, familiarize yourself with the first passage in your group in order to facilitate this process.**
6. For controversial results:
 - Encourage the panel to discuss their results with each other. Stress that the purpose of the discussion is to reach a group consensus with which everyone feels comfortable.
 - First, ask each panel member to explain the reason(s) for his or her decisions.
 - Then, facilitate a group discussion about the results. Stress that the purpose of the discussion is to reach a final group consensus.
 - After the discussion, confirm the final **group** decision.
 - If the panel has concerns about the final group consensus, have them do a secondary classification and/or add comments.
 - The facilitator can facilitate this by asking: “What are other comments about the item you want to add to this decision, or do you want to add a secondary classification for this item?”

Classification of the Remaining Items

7. First, let the panel read the passage and decide the text type and the text type subcategory.
8. Let the panel members classify all the items for the passage **individually** one at a time, and then have them check results with each other by going through each item **as a group**. You can initiate this by saying:
 - Now, you are going to classify _____ items at a time. Begin by completing the classification **individually**.
 - Then, we'll get together and discuss the results and work toward a group consensus for each item. We'll repeat this process until we've classified all of the items.
9. For conflicting results, facilitate a group discussion to reach a group consensus.
 - Keep their discussions on track (should focus on the cognitive processes of the items by the NAEP framework).
 - If the pace seems too slow, you may want to say something like: "We have a limited amount of time to talk about a single item. For the sake of the time available, we will need to wrap up our discussion on this item and try to reach a final decision as a group."
10. Monitor the time and remind the panel of the pace we will need to keep to complete all items by the end of the meeting:
 - Maximum time allowed for each passage: 120 minutes, including discussion.
 - Give about 5~10 minutes for the experts to read the passage and decide its text type. Then, give the experts about 20-30 minutes do individual classifications and save the rest of time for group discussion.
 - Let the panel have breaks and lunch on time.

Group Wrap-up Day One and Day Two (30 minutes)

11. After the item classification exercises on each day, guide the panel to discuss **points 1 through 4 in the Discussion Guide**.
 - **Discussion Point #1:** What are the differences and similarities between the two assessments in terms of:
 - Definition of reading
 - ✓ What are the similarities in how PIRLS and NAEP define reading?
 - ✓ What are the differences in what is emphasized in the definitions of reading in PIRLS and NAEP?
 - ✓ Other observations
 - ✓ Overall conclusions
 - Reading purposes and text types
 - ✓ What are the similarities and differences in the reading purposes and text type selections in PIRLS and NAEP?
 - ✓ Other observations
 - ✓ Overall conclusions

- Definition of comprehension processes and cognitive targets
 - ✓ What are the similarities and differences between the PIRLS comprehension process “focus on and retrieve explicitly stated information” and the NAEP cognitive target “locate and recall”?
 - ✓ What are the similarities and differences between the two PIRLS comprehension processes “make straightforward inferences, integrate” and “interpret ideas and information” and the NAEP cognitive target “integrate and interpret”?
 - ✓ What are the similarities and differences between the PIRLS comprehension process “examine and evaluate content, language, and textual elements” and the NAEP cognitive target “critique and evaluate”?
 - Other observations
 - Overall conclusions
 - **Discussion Point # 2:** Could the given passages from PIRLS be used in NAEP according to the text types defined in the NAEP framework? Why or why not?
 - **Discussion Point # 3:** If an item could not be mapped into any of the cognitive targets of NAEP, is there anything about the cognitive requirements of the item that would set it outside of the NAEP framework?
 - **Discussion Point # 4:** Among the items that could not be mapped into the cognitive targets of NAEP, what are the similarities among them (e.g., cognitive processes, item structures)?
 - Have the note taker read out the item numbers that have been set outside of the framework. (Those with an X in the cognitive target cell.)
 - **Discussion Point #5:** What are the characteristics of these items outside the U.S. fourth-grade curriculum?
 - Have the note taker read out the item numbers that are outside U.S. curricula.
 - Have the experts review these items for several minutes.
 - Prompt the experts to share their observations on the characteristics of these items and summarize their responses.
 - ✓ “Why are these items outside U.S. curricula?”
 - ✓ “Are there any similarities among these items, such as a similar cognitive demand?”
12. **Day One:** Close the discussion at 5 p.m. and remind the panel to come back at 8:30 a.m. in the same room to continue item classification.
13. **Day Two:** Have the panel rejoin the entire group at 4 p.m. for a plenary wrap-up.