

Chapter 4

Assessment Design and Specifications



4

Chapter 4

Assessment Design and Specifications

PIRLS 2011 consists of a wide-ranging assessment of reading comprehension to measure fourth-grade students' reading literacy achievement as well as a series of questionnaires focusing on contexts for reading literacy development to gather information about community, home, and school contexts for developing reading literacy. Conducted on a regular five-year cycle, with each assessment linked to those that preceded it, PIRLS provides regular data on trends in students' reading literacy on a common achievement scale.

By assessing students at the fourth grade, PIRLS provides data that complement TIMSS, IEA's Trends in International Mathematics and Science Study, which assesses achievement at fourth and eighth grades every four years. The fifth in the TIMSS series of assessments, TIMSS 2011 will be the first TIMSS assessment to have data collection in the same school year as PIRLS, providing a rare opportunity for countries to collect internationally comparable information on reading, mathematics, and science in the same year and on the same students.¹

Student Population Assessed

PIRLS assesses the reading literacy of children in their fourth year of formal schooling. This population was chosen for PIRLS because it is an important transition point in children's development as readers. Typically, at this point, students have learned how to read and are now reading to learn. In many countries also, this is where

¹ Countries participating in PIRLS and TIMSS at the fourth grade in 2011 will have the option of administering the assessments to the same students or to separate student samples. Most countries are planning to administer the two assessments to the same students.

students begin to have separate classes for different subjects, such as mathematics and science. The target population for PIRLS is defined as follows.

The PIRLS target grade should be the grade that represents four years of schooling, counting from the first year of ISCED Level 1.

ISCED is the International Standard Classification of Education developed by the UNESCO Institute for Statistics and provides an international standard for describing levels of schooling across countries. The ISCED system describes the full range of schooling, from preprimary (Level 0) to the second level of tertiary education (level 6). ISCED Level 1 corresponds to primary education or the first stage of basic education. The first year of Level 1 should mark the beginning of “systematic apprenticeship of reading, writing and mathematics” (UNESCO, 1999). Four years later would be the PIRLS target grade, which is the fourth grade in most countries. However, given the linguistic and cognitive demands of reading, PIRLS wants to avoid assessing very young children. Thus, PIRLS recommends that countries assess the next higher grade (i.e., fifth grade) if the average age of fourth grade students at the time of testing would be less than 9.5 years.

Reporting Reading Achievement

PIRLS 2011 will provide a comprehensive picture of the reading literacy achievement of the participating students in each country. This will include achievement by reading purpose and comprehension process as well as overall reading achievement. Consistent with the goal of a comprehensive view of reading comprehension, the complete PIRLS 2011 assessment consists of ten reading passages and accompanying questions (known as items) and, similarly, the

prePIRLS assessment consists of six less difficult reading passages and accompanying questions. However, to keep the assessment burden on any one student to a minimum, each student is presented with only part of the assessment according to a systematic booklet assembly and rotation procedure, as described in the next section. Following data collection, student responses are placed on a common reading achievement scale using item response theory methods that provide an overall picture of the assessment results for each country.²

PIRLS was designed from the outset to measure trends over time in reading achievement. Accordingly, the PIRLS reading achievement scale provides a common metric on which countries can compare their fourth grade students' progress in reading over time from assessment to assessment. The PIRLS achievement scale was established in 2001 to have a scale average of 500 and a standard deviation of 100, corresponding to the international mean and standard deviation of the countries that participated in PIRLS 2001. Using passages that were administered in both 2001 and 2006 assessments as a basis for linking the two sets of assessment results, the PIRLS 2006 data also were placed on this scale so that countries could gauge changes in students' reading achievement since 2001. A similar procedure will be employed for PIRLS 2011, so that PIRLS 2011 countries that have participated in PIRLS since its inception can have comparable achievement data from 2001, 2006, and 2011, and can plot changes in performance over this period.

To complement the overall reading literacy scale and similar to PIRLS 2006, PIRLS 2011 will provide separate achievement scales for purposes for reading and for processes of comprehension. More specifically, there will be two scales for reading purposes:

- reading for literary experience
- reading to acquire and use information.

2 The PIRLS scaling methodology is described in detail in Foy, Galia, & Li (2007).

There also will be two scales for processes of reading comprehension:

- retrieval and straightforward inferencing
- interpreting, integrating, and evaluating.³

Test Booklet Design

Given the broad coverage and reporting goals of the PIRLS framework and its emphasis on the use of a variety of authentic texts, it was inevitable that the specifications for the item pool would include extensive testing time. The PIRLS Reading Development Group found that a valid assessment of two purposes for reading, reading for literary experience and reading to acquire and use information, with reliable measures of two processes of comprehension required at least six hours of testing time. While the assessment material that can be presented in that time should provide good coverage of the reading material children meet in their everyday lives, it is not reasonable to expect to administer the entire set of reading passages and test items to any one child. Because of the difficulties of scheduling student assessments and because young children cannot be subjected to long testing periods without suffering loss of concentration and fatigue, the testing time is limited to 80 minutes per student, with an additional 15–30 minutes for a student questionnaire.

With a total testing time of more than six hours but far less than that advisable for any individual student, the assessment material must be divided among students in some way. The PIRLS design uses a matrix sampling technique, whereby the passages and accompanying items are divided into groups or blocks, and individual student booklets are made up from these blocks according to a systematic arrangement.

³ Retrieval and straightforward inferencing combines items from the Focus on and retrieve explicitly stated material and Make straightforward inferences comprehension processes. Similarly, Interpreting, integrating, and evaluating is based on items from the Interpret and integrate ideas and information and Examine and evaluate content, language, and textual elements processes.

In PIRLS 2011, similar to PIRLS 2006, the more than six hours of testing time is divided into ten 40-minute blocks of passages and items, labeled L1–L5 for the literary passages and I1–I5 for the informational texts (see Figure 3). Six of the ten blocks were included in previous PIRLS assessments—two in both PIRLS 2001 and PIRLS 2006 and four in PIRLS 2006 only. These “trend” blocks provide a foundation for measuring trends in reading achievement. Four new blocks will be developed for use for the first time in the 2011 assessment.

Figure 3 PIRLS 2011 Matrix Sampling Blocks

Purpose for Reading	Block				
Literary Experience	L1	L2	L3	L4	L5
Acquire and Use Information	I1	I2	I3	I4	I5

Similar to PIRLS 2006, the ten blocks of passages and items in the PIRLS 2011 design will be distributed across 13 booklets (see Figure 4). Each student booklet will consist of two 40-minute blocks of passages and items. Each student will respond to one assessment booklet and a student questionnaire. So as to present at least some passages in a more natural, authentic setting, two blocks (one literary and one informational) will be presented in a magazine-type format with the questions in a separate booklet. This booklet is referred to as the PIRLS “Reader.”

To enable linking among booklets, at least some blocks must be paired with others. Since the number of booklets can become very large if each block is to be paired with all other blocks, it was necessary to choose judiciously among possible block combinations. In the 13-booklet design used in PIRLS 2006, 12 test booklets are derived by combining four literary (L1, L2, L3, and L4) and four

informational (I1, I2, I3, and I4) blocks in a variety of configurations. The 13th booklet, the Reader, accounts for the remaining literary block, L5, and informational block, I5.

In this design, each of blocks L1 through L4 and I1 through I4 appear in three of the 12 booklets, each time paired with another, different, block. For example, as shown in Figure 4, literary block L1 appears with literary block L2 in Booklet 1 and with informational blocks I4 and I1 in Booklets 8 and 9. Similarly, literary block L2 appears not only with L1 in Booklet 1 but also with literary block L3 in Booklet 2 and with informational block I2 in Booklet 10.

The pairing of blocks in Booklets 1 through 12 ensures that there are good links both among the literary and among the informational passages and also between the two purposes for reading. The blocks in the Reader, L5 and I5, are not linked to any other blocks directly. However, because booklets are assigned to students using a randomized procedure, the group of students responding to the Reader is equivalent to those responding to the other booklets, within the margin of error of the sampling process. Because each block appears in three of Booklets 1 through 12, the Reader is assigned three times more frequently in the distribution procedure than these Booklets so that the same proportion of students respond to blocks L5 and I5 as to each of the other literary and informational blocks.

Selecting Reading Passages for the Assessment

To reach the goal of approximating an authentic reading experience in the assessment, the reading passages presented to students must be typical of those read by students in their everyday experiences. Texts that exist for students to read in and outside school have typically been written by successful authors who understand writing for a young audience. These are more likely than passages written specifically for a test to elicit the full range of comprehension processes.

Figure 4 PIRLS 2011 Student Booklet Design

Booklet	Part 1	Part 2
1	L1	L2
2	L2	L3
3	L3	L4
4	L4	I1
5	I1	I2
6	I2	I3
7	I3	I4
8	I4	L1
9	L1	I1
10	I2	L2
11	L3	I3
12	I4	L4
Reader	L5	I5

Furthermore, they are more likely to engage students' interests, and to yield assessment questions that will elicit a range of responses to text that are similar to those elicited in authentic reading experiences. In the context of an international study, attaining authenticity in the assessment reading experience may be somewhat constrained by the need to translate a text into numerous languages. Thus, care is taken to choose texts that can be translated without loss in meaning or in potential for student engagement.

In selecting texts for use in an international survey of reading literacy, the potential for cultural bias must be considered. The set

of texts used must range as widely as possible across nations and cultures. No country or culture should be overrepresented in the assessment texts. Text selection thus involves collecting potential stimulus texts from as many countries as possible. The final selection of texts is based, in part, on the national and cultural representation of the entire set of assessment texts. Texts that depend heavily on culture-specific knowledge are excluded.

The appropriateness and readability of texts for assessing fourth-grade students is determined through review by educators and curriculum specialists from countries participating in the assessment. Among the criteria used to select texts are topic and theme appropriateness for the grade level; fairness and sensitivity to gender, racial, ethnic, and religious considerations; nature and level of linguistic features; and density of information. In addition, the time constraints of the test situation place some limits on the length of texts. Generally, texts selected for 2011 will be no longer than 800 words so students have time to read the entire passage and answer the comprehension questions. However, length will vary somewhat because other text characteristics also affect rate of reading.

As a basis for measuring trends from 2001 and 2006, PIRLS retained six passages and items from the 2001 and 2006 assessments—three literary and three informational—to be included in the PIRLS 2011 assessment. To complete the design, four new passages and associated items will be selected—two literary and two informational.

Question Types and Scoring Procedures

Students' ability to comprehend text through the four PIRLS comprehension processes is assessed via comprehension questions that accompany each text. Two question formats are used in the PIRLS assessment—multiple-choice and constructed-response.

Each multiple-choice question is worth one point. Constructed-response questions are worth one, two, or three points, depending on the depth of understanding required. Up to half of the total number of points represented by all of the questions will come from multiple-choice questions. In the development of comprehension questions, the decision to use either a multiple-choice or a constructed-response format is based on the process being assessed, and on which format best enables test takers to demonstrate their reading comprehension.

Multiple-Choice Questions

Multiple-choice questions provide students with four response options, of which only one is correct. Multiple-choice questions can be used to assess any of the comprehension processes. However, because they do not allow for students' explanations or supporting statements, they may be less suitable for assessing students' ability to make more complex interpretations or evaluations.

In assessing fourth-grade students, it is important that linguistic features of the questions be developmentally appropriate. Therefore, the questions are written clearly and concisely. The response options are also written succinctly in order to minimize the reading load of the question. The options that are incorrect are written to be plausible, but not deceptive. For students who may be unfamiliar with this test question format, the instructions given at the beginning of the test include a sample multiple-choice item that illustrates how to select and mark an answer.

Constructed-Response Questions

For this type of test item students are required to construct a written response, rather than select a response from a set of options. The emphasis placed on constructed-response questions in the PIRLS assessment is consistent with the definition of literacy underlying the framework.

It reflects the interactive, constructive view of reading—meaning is constructed through an interaction between the reader, the text, and the context of the reading task. This question type is used to assess any of the four comprehension processes. However, it is particularly well suited for assessing aspects of comprehension that require students to provide support or that result in interpretations depending upon students’ background knowledge and experiences.

In the PIRLS assessment, constructed-response questions may be worth one, two, or three points, depending on the depth of understanding or the extent of textual support the question requires. In framing these questions, it is important to provide enough information to help students understand clearly the nature of the response expected.

Each constructed-response question has an accompanying scoring guide that describes the essential features of appropriate and complete responses. Scoring guides focus on evidence of the type of comprehension the questions assess. They describe evidence of partial understanding and evidence of complete or extensive understanding. In addition, sample student responses at each level of understanding provide important guidance to scoring staff.

In scoring students’ responses to constructed-response questions, the focus is solely on students’ understanding of the text, not on their ability to write well. Also, scoring takes into account the possibility of various interpretations that may be acceptable, given appropriate textual support. Consequently, a wide range of answers and writing ability may appear in the responses that receive full credit to any one question.

Score Points

In developing the assessment, the aim is to create blocks that each provide, on average, at least 15 score points—made up of

approximately seven multiple-choice items (1 point each), two or three short-answer items (1 or 2 points each), and one extended-response item (3 points). Items in each block should address the full range of PIRLS comprehension processes. The exact number of score points and the exact distribution of question types per block will vary somewhat, as different texts yield different types of questions.

Releasing Assessment Material to the Public

An essential aspect of the PIRLS design for measuring trends over time in reading achievement is that, with each cycle, PIRLS releases a number of passages and items into the public domain to help readers understand as much as possible about the content and approach of the assessment. At the same time, a number of passages and items are retained and kept confidential to be used in future assessments as the basis for measuring trends. As passages and items are released, new assessment material is developed to take their place.

According to the PIRLS design, four blocks were released following the PIRLS 2006 data collection, two from the original 2001 assessment (L2 and I2) and two from those developed for 2006 (L5 and I5). These released passages and items may be found in the *PIRLS 2006 International Report* (Mullis, Martin, Kennedy, & Foy, 2007). Following the publication of the international report for PIRLS 2011, a further four blocks will be released, two that were used in both the 2006 and 2011 assessments and two from those developed for specifically for PIRLS 2011.

prePIRLS 2011 Assessment Design

The prePIRLS assessment follows the PIRLS student populations and assessment design as closely as possible, with changes made only when appropriate for the prePIRLS context. Also similar to PIRLS, prePIRLS provides a broad picture of students' reading achievement by placing student responses on a common scale and measuring trends in achievement over time.

Like PIRLS, prePIRLS will report student achievement on an overall reading literacy scale and separate achievement scales for the literary and informational reading purposes. The reading comprehension process scales, however, differ from PIRLS due to the greater emphasis on the process of focusing on and retrieving explicitly stated information in the prePIRLS items. More specifically, the two scales for the processes of comprehension will be:

- retrieving explicitly stated information
- inferencing, interpreting, and evaluating.⁴

Following the design that has proven effective in previous cycles of PIRLS, prePIRLS also uses a matrix sampling technique for the administration of the assessment. PrePIRLS is comprised of six blocks of reading passages and their accompanying items, for a total of four hours of testing time (see Figure 5). Because prePIRLS is newly developed for 2011, all of these test blocks will be new and four of the test blocks will be kept secure to measure trends in future assessment cycles.

Figure 5 prePIRLS 2011 Matrix Sampling Blocks

Purpose for Reading	Block		
Literary Experience	L1	L2	L3
Acquire and Use Information	I1	I2	I3

These six test blocks are distributed across nine student booklets, with each block appearing in three booklets to enable linking between the various blocks (see Figure 6). As in PIRLS, each booklet contains

4 Retrieving explicitly stated information uses items from the Focus on and retrieve explicitly stated material comprehension process. Inferencing, interpreting, and evaluating is based on items from the Make straightforward inferences, Interpret and integrate ideas and information, and Examine and evaluate content, language, and textual elements processes.

Figure 6 prePIRLS 2011 Student Booklet Design

Booklet	Part 1	Part 2
1	L1	L2
2	L2	L3
3	L3	I1
4	I1	I2
5	I2	I3
6	I3	L1
7	L1	I1
8	I2	L2
9	I3	L3

two 40-minute test blocks, for a total of 80 minutes of testing time per student followed by 15-30 minutes for a student questionnaire.

Like PIRLS, prePIRLS strives to replicate an authentic reading experience with materials that are engaging and familiar to students participating in prePIRLS. However, it is sometimes necessary to adapt texts to ensure that the subject matter is age-appropriate, while keeping the reading load reasonable for prePIRLS students. Passages selected for prePIRLS are generally no longer than 400 words in length to ensure that students have ample time to read the passage and respond to the accompanying items. As an additional step to help students locate information within the text, items are interspersed throughout the passage. When possible, items that require students to focus on a particular page of text are placed on the facing page, so that students can view both the items and the relevant text simultaneously. This distribution of items also helps to

ensure that students can provide answers to some questions, even if they do not complete the entire passage.

The prePIRLS items use multiple-choice and constructed-response formats, as in PIRLS, though constructed-response items are worth only one or two points. However, there is a slightly higher percentage of constructed-response items in the prePIRLS assessment, comprising up to 60 percent of the total score points. This decision was made because constructed-response items that require a very short response are often easier for early readers due to the lighter reading load, as compared with multiple-choice items that require students to read and evaluate response four options. In addition, multiple-choice items may lose some of their effectiveness in passages as short as those used in prePIRLS, as there are fewer plausible distracters that can be drawn from the text.

Background Questionnaires

An important purpose of PIRLS and prePIRLS is to study the community, home, and school factors associated with children's reading literacy by the fourth grade. To that end, PIRLS and prePIRLS 2011 will administer questionnaires to students, their parents, their teachers, and the principals of their schools. The questions are designed to measure key aspects of students' home and school environments. PIRLS and prePIRLS 2011 also will request the National Research Coordinators to complete a curriculum questionnaires for their countries.

Student Questionnaire

A questionnaire will be completed by each student who participates in the PIRLS reading assessment. It asks about aspects of students' home and school lives, including demographic information, home environment, school climate for learning, out-of-school reading behaviors, and attitudes toward reading. The student questionnaire requires 15–30 minutes to complete.

Learning to Read Survey (Home Questionnaire)

This short questionnaire is addressed to the parents or primary caregivers of each student taking part in the PIRLS 2011 data collection. It asks about language spoken in the home, preschool experiences, homework activities, home-school involvement, books in the home, and parents' education and involvement. Also, it collects information on early literacy and numeracy activities, reading and quantitative readiness, parents' reading activities and attitudes toward reading. Together with information collected from the students, parents' responses will provide a more complete picture of an important context for learning to read. This questionnaire is designed to take 10–15 minutes to complete.

Teacher Questionnaire

The reading teacher of each fourth-grade class in PIRLS 2011 will be asked to complete this questionnaire, which is designed to gather information about teacher characteristics and classroom contexts for developing reading literacy. The questionnaire asks teachers about their background and education, the school climate for learning, attitudes toward teaching, classroom characteristics, and student engagement. It also asks about reading instructional time, approaches, activities, and materials; computer and library resources; homework; and preparation to teach reading. This questionnaire requires about 30 minutes of the teacher's time.

School Questionnaire

The principal of each school in PIRLS 2011 will be asked to respond to this questionnaire. It asks about school characteristics, instructional time, resources and technology, parental involvement, school climate for learning, teaching staff, the role of the principal, and students' reading readiness. It is designed to take about 30 minutes.

Curriculum Questionnaire

To provide information about the goals of reading instruction, the national research coordinator in each country will complete a questionnaire about the country's reading curriculum, including national policy on reading, goals and standards for reading instruction, time specified for reading, and provision of books and other literary resources.

PIRLS 2011 Encyclopedia

The PIRLS 2011 Encyclopedia will provide a profile of each country's education system, with a particular focus on reading education for primary-school children. The volume will provide general data on economic and educational indicators and describe how the education system is organized and how decisions about education are made. The reading curriculum, including goals, materials, and instruction, will be discussed, along with information on assessment of reading achievement.