

CHAPTER 3

Assessment Design for PIRLS, PIRLS Literacy, and ePIRLS in 2016

Michael O. Martin, Ina V.S. Mullis, and Pierre Foy

PIRLS 2016 consists of three separate assessments of reading comprehension: PIRLS, PIRLS Literacy, and ePIRLS. PIRLS is a comprehensive assessment of fourth grade students' reading literacy achievement. 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. Matching PIRLS for breadth of coverage but with less difficult reading passages and items, PIRLS Literacy extends the effective measurement of reading comprehension at the lower end of the PIRLS achievement scale. For countries participating in PIRLS, ePIRLS expands PIRLS to include the assessment of online reading to acquire and use information. The PIRLS assessments include a series of contextual questionnaires to gather information about community, home, and school contexts for developing reading literacy.

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, this also is when 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 (UNESCO, 2012). The ISCED system describes the full range of schooling, from early childhood education (Level 0) to doctoral study (Level 8). ISCED Level 1 corresponds to primary education, or the first stage of basic education. The PIRLS target grade is four years after the beginning of Level 1, 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, if the average age of fourth grade students at the time of testing would be less than 9.5 years, PIRLS recommends that countries assess the next higher grade (i.e., fifth grade).

Reporting Reading Achievement

PIRLS and PIRLS Literacy are designed to provide a complete picture of the reading literacy achievement of the participating students in each country. This includes 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 entire PIRLS assessment consists of 12 reading passages and accompanying questions (known as items); similarly, the PIRLS Literacy assessment consists of 12 reading passages and accompanying questions, but the passages are less difficult. In each assessment, six passages assess reading for literary experience and six assess reading to acquire and use information. In order to keep the assessment burden on any one student to a minimum, each student is presented with just two passages according to a systematic booklet assembly and rotation procedure, as described in the next section. Following data collection, student responses for both the PIRLS and PIRLS Literacy assessments are placed on the PIRLS reading achievement scale using item response theory methods that provide an overall picture of the assessment results for each country.¹

Integration between PIRLS and PIRLS Literacy is maintained by including two PIRLS Literacy passages in the PIRLS assessment and two PIRLS passages in the PIRLS Literacy assessment. This provides a solid foundation for employing the PIRLS scaling and linking methodology to ensure that students taking the PIRLS Literacy assessment have their achievement reported on the PIRLS scale. Moreover, including the two less difficult PIRLS Literacy passages benefits PIRLS by providing more information about the reading accomplishments of students who participate in the PIRLS assessment and perform at the lower end of the

¹ The PIRLS scaling methodology is described in detail in Foy, Brossman & Galia (2012).

achievement scale. Conversely, including the more difficult PIRLS passages in the PIRLS Literacy assessment provides information about the accomplishments of higher performing students who participate in PIRLS Literacy.

The PIRLS assessments are 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 so that 100 points on the scale corresponded to one standard deviation across all of the countries that participated in 2001, and the scale centerpoint of 500 corresponded to the international average across those countries. Using passages that were administered in both the 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. Following a similar procedure, the PIRLS 2011 data also were placed on the PIRLS scale, as will be the data from PIRLS 2016. This will enable countries that have participated in PIRLS since its inception to have comparable achievement data from 2001, 2006, 2011, and 2016, and to plot changes in performance over this 15-year period.

The PIRLS reading achievement scale is an overall measure of reading proficiency that includes both reading purposes and processes of comprehension. However, in addition to the overall scale, PIRLS and PIRLS Literacy also provide separate achievement scales on the same metric for purposes for reading and for processes of comprehension. More specifically, there are two scales for reading purposes:

- Reading for literary experience; and
- Reading to acquire and use information.

In addition to these, there also are two scales for processes of reading comprehension:

- Retrieval and straightforward inferencing; and
- Interpreting, integrating, and evaluating.²

Countries participating in ePIRLS also participate in PIRLS; so, in addition to the usual PIRLS overall reading achievement results and results by reading purpose and comprehension process, ePIRLS participants can report student

2 Retrieval and straightforward inferencing combines items from the Focus on and Retrieve Explicitly Stated Information 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 Critique Content and Textual Elements processes.

achievement in online reading for informational purposes. The ePIRLS online reading achievement scale enables countries to examine their students' online reading performance relative to their performance on the PIRLS reading achievement scales.

PIRLS and PIRLS Literacy 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, the specifications for the pool of assessment items 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 good coverage of the range of reading material that children encounter in school and their everyday lives.

With a total testing time for the assessment passages of eight hours, but far less time available to assess any individual student, the PIRLS assessment materials must be divided in some way. Therefore, 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.

To address this challenge, the PIRLS assessment design uses a matrix sampling technique: each reading passage and its accompanying items is assigned to a block, and the blocks are then systematically distributed among individual student booklets. Both PIRLS and PIRLS Literacy consist of 12 passages/blocks, each of which is expected to require 40 minutes of student testing time.

As shown in Exhibit 3, the five literary blocks developed specifically for PIRLS are labeled PRLit1 through PRLit5 and the five informational blocks PRInf1 through PRInf5. The two blocks from PIRLS Literacy are labeled PLLit3 and PLInf3. Six of the ten PIRLS blocks were included in previous PIRLS assessments: two in all three assessments (2001, 2006, and 2011), two in both PIRLS 2006 and PIRLS 2011, and two in PIRLS 2011 only. These “trend” blocks provide a foundation for measuring trends in reading achievement. In addition, the 2016 assessment includes four new PIRLS blocks developed for use for the first time.

Exhibit 3: PIRLS 2016 Matrix Sampling Blocks

Purpose for Reading	Block					
Literary Experience	PRLit1	PRLit2	PRLit3	PRLit4	PRLit5	PLLit3
Acquire and Use Information	PRInf1	PRInf2	PRInf3	PRInf4	PRInf5	PLInf3

The ten blocks developed specifically for PIRLS Literacy are shown in Exhibit 4, with the five blocks of literary passages labeled PLLit1 through PLLit5 and the five informational blocks PLInf1 through PLInf5. The two blocks from PIRLS are labeled PRLit1 and PRInf1. Four of the passage and item blocks were previously used in 2011 as part of prePIRLS. Because prePIRLS has been subsumed into PIRLS Literacy for the 2016 assessment cycle, these passages from 2011 provide the basis for measuring trends in 2016. The remaining six PIRLS Literacy blocks are newly developed for 2016.

Exhibit 4: PIRLS Literacy 2016 Matrix Sampling Blocks

Purpose for Reading	Block					
Literary Experience	PLLit1	PLLit2	PLLit3	PLLit4	PLLit5	PRLit1
Acquire and Use Information	PLInf1	PLInf2	PLInf3	PLInf4	PLInf5	PRInf1

The PIRLS 2016 booklet design shows how the blocks of passages and items are assembled into individual student booklets, each consisting of two 40-minute blocks of passages and items. Individual students respond to one assessment booklet and a student questionnaire.

The PIRLS booklet design (see Exhibit 5) includes the ten blocks of PIRLS passages and items described in Exhibit 3, as well as two of the PIRLS Literacy blocks from Exhibit 4 (PLLit3 and PLInf3). These 12 blocks are distributed across 16 booklets. Booklets 1–15 each consist of one literary passage and items and one informational passage and items. In order to present at least some passages in a more natural, authentic setting, one literary block (PRLit5) and one informational block (PRInf5) are presented in a magazine-type format with the questions in a separate booklet. This 16th booklet is referred to as the PIRLS “Reader.”

Exhibit 5: PIRLS 2016 Student Booklet Design

Booklet	Part 1	Part 2
1	PRInf2	PRLit1
2	PRLit3	PRInf2
3	PLInf3	PRLit3
4	PLInf3	PRLit4
5	PRLit4	PRInf1
6	PRLit2	PRInf1
7	PRInf3	PRLit2
8	PLLit3	PRInf3
9	PLLit3	PRInf4
10	PRInf4	PRLit1
11	PRLit3	PRInf1
12	PLInf3	PRLit2
13	PRInf3	PRLit1
14	PLLit3	PRInf2
15	PRInf4	PRLit4
Reader	PRLit5	PRInf5

The 16 PIRLS booklets are distributed among students in participating classrooms so that the groups of students completing each booklet are approximately equivalent in terms of student reading ability. PIRLS uses item response theory scaling methods to assemble a comprehensive picture of the reading achievement of a country’s entire fourth grade student population by pooling individual students’ responses to the booklets that they are assigned. This approach reduces to manageable proportions what otherwise would be an impossible student burden, albeit at the cost of greater complexity in booklet assembly, data collection, and data analysis.

In order to enable linking among booklets within PIRLS, and to maintain links between PIRLS and PIRLS Literacy, it is desirable that the student booklets contain as many block pair combinations as possible. However, because the number of booklets can become very large if each block is to be paired with all other blocks, it is necessary to choose judiciously among possible block combinations.

In the PIRLS 16-booklet design, each of five literary blocks (PRLit1–PRLit4 and PLLit3) and each of five informational blocks (PRInf1–PRInf4 and PLInf3) appear in three of the PIRLS booklets, each time paired with another, different block. For example, as shown in Exhibit 5, literary block PRLit1 appears with informational block PRInf2 in Booklet 1 and with informational blocks PRInf4 and PRInf3 in Booklets 10 and 13, respectively. Informational block PRInf2 appears not only with PRLit1 in Booklet 1, but also with literary block PRLit3 in Booklet 2 and with PIRLS Literacy literary block PLLit3 in Booklet 14. Each of the two PIRLS Literacy blocks (PLLit3 and PLInf3) appears in three PIRLS booklets. By design, the two PIRLS Literacy block passages are less demanding than the PIRLS passages. Accordingly, when a PIRLS Literacy block is paired with a PIRLS block the Literacy block always is in first position in the booklet. Including the two PIRLS Literacy blocks in the PIRLS booklet scheme ensures a link between PIRLS and PIRLS Literacy. This link is further strengthened by including two PIRLS blocks in the PIRLS Literacy booklet scheme (see below).

The blocks in the PIRLS Reader, PRLit5 and PRInf5, 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 15, 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 PRLit5 and PRInf5 as to each of the other blocks in the PIRLS booklets.

Similar to the PIRLS booklet design, the PIRLS Literacy booklet design consists of Booklets 1–15 and a PIRLS Literacy Reader, with each booklet consisting of two 40-minute blocks of passages and items, and each student responding to one assessment booklet and a student questionnaire (see Exhibit 6). Each booklet contains one literary passage and one informational passage. The PIRLS Literacy design includes the ten blocks of PIRLS Literacy passages and items shown in Exhibit 4 (PLLit1–PLLit5 and PLInf1–PLInf5) together with two of the PIRLS blocks from Exhibit 3 (PRLit1 and PRInf1). The PIRLS Literacy Reader consists of literary block PLLit5 and informational block PLInf5.

Exhibit 6: PIRLS Literacy 2016 Student Booklet Design

Booklet	Part 1	Part 2
1	PLInf1	PRLit1
2	PLLit2	PLInf1
3	PLInf3	PLLit2
4	PLInf3	PLLit4
5	PLLit4	PRInf1
6	PLLit1	PRInf1
7	PLInf2	PLLit1
8	PLLit3	PLInf2
9	PLLit3	PLInf4
10	PLInf4	PRLit1
11	PLLit2	PRInf1
12	PLInf3	PLLit1
13	PLInf2	PRLit1
14	PLLit3	PLInf1
15	PLInf4	PLLit4
Reader	PLLit5	PLInf5

Also paralleling the PIRLS design, each of five literary blocks (PLLit1–PLLit4 and PRLit1) and five informational blocks (PLInf1–PLInf4 and PRInf1) appears in three of the 15 PIRLS Literacy booklets, each time paired with another, different, block. Each of the two PIRLS blocks (PRLit1 and PRInf1) appears in three PIRLS Literacy booklets. Because these PIRLS blocks are more difficult than the PIRLS Literacy blocks, they appear in the second position in the booklet when paired with a Literacy block.

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 and PIRLS Literacy assessments: 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 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, multiple-choice questions 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, questions are written clearly and concisely. Response options also are written succinctly in order to minimize the reading demand of the question. Incorrect options 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

Constructed-response test items require students to provide a written response, rather than select a response from a set of options. The emphasis placed on constructed-response questions in the PIRLS assessments 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 may be 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 involving students' background knowledge and experiences.

In the PIRLS assessments, constructed-response questions may be worth one or two points (short-answer items), or three points (extended-response items), 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 clearly understand 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. The guides 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 PIRLS and PIRLS Literacy assessments, the aim is to create blocks of passages and items that each provide, on average, at least 15 score points consisting of the following: 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, because different texts yield different types of questions.

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

Releasing Assessment Materials to the Public

An essential aspect of the PIRLS design for measuring trends in reading achievement over time is that, with each cycle, PIRLS releases a number of passages and items into the public domain in order 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 materials are developed to take their place.

According to the PIRLS design, four blocks were released following the PIRLS 2011 data collection, two developed originally for the 2006 assessment, and two from the four developed for 2011. These released passages and items may be found in the *PIRLS 2011 International Results in Reading* (Mullis, Martin, Foy, & Drucker, 2012). Following the publication of the international report for PIRLS 2016, a further six blocks will be released: four that were used in both the 2011 and 2016 assessments, and two from those developed specifically for PIRLS 2016. Additionally, the two PIRLS passages that were included in the PIRLS Literacy booklet design will be released, along with two PIRLS Literacy blocks from 2011 and two from 2016.

ePIRLS 2016 Design

The ePIRLS computer-based assessment of online reading is designed as an extension to PIRLS that measures student informational reading in an online environment. ePIRLS is administered by computer, and requires students to use a mouse or other pointing device to navigate through the assessment and to use a computer keyboard to type their responses to the assessment questions. All students participating in ePIRLS also are expected to have participated in PIRLS. The complete ePIRLS assessment consists of four³ school-based online reading tasks, each of which involves 2–3 different websites totaling 5 to 10 web pages, together with a series of comprehension questions based on the task. Similar to the PIRLS and PIRLS Literacy passages, each task with accompanying questions takes 40 minutes to complete. In order to keep student response burden to a reasonable level, each individual student completes just two ePIRLS tasks, followed by 5 minutes for a short online questionnaire.

Because ePIRLS is administered by computer, it has greater flexibility than paper-based PIRLS in how the assessment tasks are paired for presentation to

3 Depending on the results of the ePIRLS field test, the number of assessment tasks may be increased to five or six. In that case, the matrix sampling design for task combinations will be extended. In general, if there are n tasks, the number of task combinations is n^2-n .



students. With each student taking two of the four assessment tasks, there are 12 possible task combinations based on task pair and order of administration (see Exhibit 7). ePIRLS uses IEA's WinW3S sampling software to randomly distribute all 12 task combinations across participating students so that approximately 1/12 of the student sample in each country responds to each task combination and these groups of students are approximately equivalent in terms of student ability.

Exhibit 7: ePIRLS 2016 Student Task Combinations—4 Tasks

Student Task Combination	First Task	Second Task
Task Combination #1	E01	E02
Task Combination #2	E01	E03
Task Combination #3	E01	E04
Task Combination #4	E02	E01
Task Combination #5	E02	E03
Task Combination #6	E02	E04
Task Combination #7	E03	E01
Task Combination #8	E03	E02
Task Combination #9	E03	E04
Task Combination #10	E04	E01
Task Combination #11	E04	E02
Task Combination #12	E04	E03

ePIRLS uses item response theory scaling methods to assemble a comprehensive picture of the online informational reading achievement of a country's fourth grade student population by pooling individual students' responses to the tasks that they have been assigned.

Because 2016 is the inaugural year for ePIRLS, all tasks are newly developed. After the 2016 assessment, two of the tasks will be released to the public and the remainder kept secure in order to measure trends in future ePIRLS assessment cycles.

Context Questionnaires and the PIRLS 2016 Encyclopedia

An important purpose of PIRLS 2016 is to study the home, community, school, and student factors associated with children’s reading literacy at the fourth grade. To accomplish this purpose, data about the contexts for learning to read are collected through questionnaires completed by students, as well as their parents, teachers, and principals. In addition, National Research Coordinators provide information on the national and community contexts for learning through the curriculum questionnaire and their country’s entry in the *PIRLS 2016 Encyclopedia*. Because PIRLS and PIRLS Literacy are reported together in order to assess students in their fourth year of schooling, the same set of questionnaires is used for all students.

PIRLS focuses on policy relevant topics that are generally considered to have a positive relationship with student achievement. Chapter 2 provides an overview of these topics and serves as the basis for item development. Many of the topics are measured through the use of scales—sets of items that measure the same construct. For purposes of reporting, scales are preferable over stand-alone items because they are generally more reliable and more suitable for trend measurement. For PIRLS 2011, 19 scales were reported using context questionnaire data, ranging from measures of parental attitude toward reading to measures of school climate.

Learning to Read Survey (Home Questionnaire)

The Home Questionnaire, entitled the *Learning to Read Survey*, is addressed to the parents or primary caregivers of each student taking part in the PIRLS 2016 data collection. This short questionnaire solicits information on the home context, such as languages spoken in the home, parents’ reading activities and attitudes toward reading, and parents’ education and occupation. The questionnaire also collects data on the students’ educational activities and experiences outside of school including early childhood education, early literacy and numeracy activities, and the child’s reading readiness at the beginning of primary school. This questionnaire is designed to take 10–15 minutes to complete.

Teacher Questionnaire

Students' reading teachers are asked to complete this questionnaire, which is designed to gather information about classroom contexts for reading instruction, such as characteristics of the class, reading instructional time, and instructional approaches. The questionnaire also asks about teacher characteristics, such as their career satisfaction, education, and recent professional development activities. This questionnaire requires about 35 minutes to complete.

School Questionnaire

The principal of each school is asked about school characteristics, such as student demographics, the school environment, and the availability of school resources and technology. The questionnaire also includes items focusing on the principal's leadership role, education, and experience. It is designed to take about 30 minutes.

Student Questionnaire

This questionnaire, given to each student once they have completed the reading assessment, collects information on students' home environment, such as languages spoken at home, books in the home, and other home resources for learning. This questionnaire also gathers information on student experiences in school, including feelings of school belonging and whether they are victims of bullying. Finally, the student questionnaire gathers data on out-of-school reading habits and attitudes toward reading, including whether they like reading, their confidence in reading, and their engagement in reading lessons. The student questionnaire requires 15–30 minutes to complete.

ePIRLS Student Questionnaire

In addition to the four questionnaires listed above, students also participating in ePIRLS complete a brief questionnaire as part of this computer-based assessment. The questionnaire asks students about their level of competency and experience using computers and finding information on the Internet. This questionnaire requires 5 minutes to complete.

PIRLS Encyclopedia

The *PIRLS 2016 Encyclopedia* profiles each country's education system, with a particular focus on reading education for primary school children. Each chapter provides an overview of the language/reading curriculum, as well as information on reading instruction in the primary grades. Each chapter also

includes information on the languages of instruction, teachers and teacher education, how the education system is organized, and assessment policies. In addition, each chapter provides information on the impact and use of PIRLS in the respective country.

Curriculum Questionnaire

The curriculum questionnaire complements the *PIRLS 2016 Encyclopedia* entries by collecting information from all countries about their national policies on reading curricula, goals and standards for reading instruction, and time specified for reading instruction, as well as information on preprimary education and teacher education policies.