



Chapter 2

Developing the PIRLS 2006 Reading Assessment and Scoring Guides

Ann M. Kennedy and Marian Sainsbury

2.1 Overview

Development of the PIRLS 2006 reading assessment began early in 2003 and continued until August 2005, when the international version of the assessment was finalized for data collection. The development was a collaborative process involving the PIRLS Reading Coordinator, the Reading Development Group, the PIRLS 2006 National Research Coordinators (NRCs) from the participating countries, the PIRLS Item Development Task Force, and TIMSS & PIRLS International Study Center staff.¹ The *PIRLS 2006 Framework and Specifications* (Mullis, Kennedy, Martin, & Sainsbury, 2006) provided the foundation for the assessment.

PIRLS 2006 was the second cycle of PIRLS, and was structured in a way that included new material that had a recognizable continuity with the previous test. In order to measure trends, the assessment was composed of passages and questions from PIRLS 2001, as well as new passages and items. The main purpose of the development process detailed here was to identify new passages and develop the accompanying items in a way that would continue and expand the range of the assessment model established in 2001. A timeline of the test development process is provided in Exhibit 2.1.

The NRCs were responsible for submitting and approving reading passages for the assessment and were directly involved in developing test items and scoring

¹ Marian Sainsbury of the National Foundation for Educational Research in England served as the PIRLS Reading Coordinator. The Item Development Task Force included the PIRLS Reading Coordinator, staff from the TIMSS & PIRLS International Study Center, and Patricia Donahue of Educational Testing Service. Members of the Reading Development Group, National Research Coordinators, and TIMSS & PIRLS International Study Center staff are acknowledged in Appendix A.

guides for constructed-response items. A field test was conducted in March–April 2005 that provided information about the measurement properties of potential passages and items across the countries. Based on the field-test results, the passages and items were selected and finalized for main data collection.

Exhibit 2.1 Overview of the Test Development Process

Date	Group and Activity
March 2003	TIMSS & PIRLS International Study staff begins initial search for PIRLS 2006 passages and sends a call for passages to National Research Coordinators.
September 2003	National Research Coordinators recommend updates to the framework and begin passage review.
January 2004	Reading Development Group reviews draft PIRLS assessment framework, reviews passages and recommends initial passage pool for field-test item development, and reviews draft item and scoring guide development manual.
February 2004	National Research Coordinators give final approval of the PIRLS assessment framework, select final passages for field-test item development, and participate in an item and scoring guide development workshop.
August 2004	Reading Development Group reviews field-test item pool and scoring guides.
November 2004	National Research Coordinators finalize selection of field-test item pool and scoring guides.
March 2005	National Research Coordinators are trained in applying field-test scoring guides for constructed-response items.
March-April 2005	PIRLS 2006 field test is administered.
July 2005	Reading Development Group reviews field-test results and recommends selection for main data collection.
August 2005	National Research Coordinators review field-test results and select operational passages and items.
October-December 2005	PIRLS 2006 data collection is conducted in Southern Hemisphere countries.
November 2005	Southern Hemisphere National Research Coordinators are trained in applying scoring guides for constructed-response items.
March 2006	Northern Hemisphere National Research Coordinators are trained in applying scoring guides for constructed-response items.
March-June 2006	PIRLS 2006 data collection is conducted in Northern Hemisphere countries.

2.2 Updating the PIRLS 2006 Assessment Framework

The PIRLS 2006 assessment framework was based on the *PIRLS 2001 Framework and Specifications* (Campbell, Kelly, Mullis, Martin, & Sainsbury, 2001). The TIMSS & PIRLS International Study Center initiated discussions of any necessary updates for PIRLS 2006 of the PIRLS 2001 framework among the NRCs and PIRLS Reading Development Group. These updates to the framework

were intended to reflect the findings from PIRLS 2001, as well as current reading research since the development of the initial framework. The process began with a review of the existing framework at the first meeting of the NRCs in September 2003, resulting in minor amendments to the definition of reading literacy and expanding the discussion and description of text types used in the assessment. NRCs also suggested an extended discussion of the interaction between purpose and text type.

The PIRLS Reading Development Group met in January 2004 to review the framework in light of the recommendations from the NRCs. In general, there was a consensus among the groups. The definition of reading literacy was reworded to underscore the importance of the variety of contexts in which reading takes place. Further adaptations included elaboration of the terms “read to learn” and “communities of readers” from the definition. The definition follows.

For PIRLS, reading literacy is defined as the 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, Kennedy, Martin, & Sainsbury, 2006, p. 3)

Modifications to the framework’s references were implemented in February 2004 following advice from the NRCs and Reading Development Group. The TIMSS & PIRLS International Study Center conducted a literature search of research articles, reports, and papers published since the publication of the 2001 framework that were relevant to the purposes for reading, processes of reading comprehension, and contexts for learning to read. A separate reference section was added to highlight reading research conducted using data from IEA studies.

The PIRLS 2006 framework was initially published in 2004, prior to the administration of the PIRLS 2006 field test. A second edition was published in February 2006 to present example reading test blocks (passages and corresponding questions) that represented the set of test blocks in the 2006 assessment. For this purpose, the appendix containing examples of two reading passages and their corresponding questions was updated with examples from the 2006 field test. The appendix also includes scoring guides for constructed-response questions.

2.3 The PIRLS 2006 Assessment Framework

The *PIRLS 2006 Assessment Framework and Specifications* contains a detailed description of the PIRLS 2006 assessment of reading comprehension. In brief, the PIRLS 2006 framework defines the two major aspects of students' reading literacy—purposes for reading and processes of comprehension. Reading for literary experience and reading to acquire and use information are the two major purposes that account for the majority of reading experiences of young children. Readers make meaning of texts in a variety of ways, depending not only on the purpose for reading, but also on the difficulty of the text and the reader's prior knowledge. PIRLS looks at four processes of comprehension: focus on and retrieve explicitly stated information; make straightforward inferences; interpret and integrate ideas and information; and examine and evaluate content, language, and textual elements. These processes are the basis for developing comprehension questions in the reading assessment. Exhibit 2.2 shows the target percentages of the reading assessment devoted to reading purposes and processes.

Exhibit 2.2 Percentages of Reading Assessment Devoted to Reading Purposes and Processes

Purposes for Reading	
Literary Experience	50%
Acquire and Use Information	50%
Processes of Comprehension	
Focus on and Retrieve Explicitly Stated Information	20%
Make Straightforward Inferences	30%
Interpret and Integrate Ideas and Information	30%
Examine and Evaluate Content, Language, and Textual Elements	20%

2.4 PIRLS 2006 Assessment Design

The PIRLS 2006 assessment design, also elaborated in the *PIRLS 2006 Framework and Specifications*, builds on PIRLS 2001, in which there were four literary and four informational test blocks. Several factors influenced the test booklet design used for PIRLS 2006 data collection. However, based on research analyses conducted by Germany and the TIMSS & PIRLS International Study Center using the PIRLS 2001 data,² NRCs requested that scaling of the PIRLS assessment be done for processes of comprehension, as well as by purposes

2 Bos, W., Lankes, E.M., Prenzel, M., Schwippert, K., Walther, G., & Valtin, R. (Hrsg.). (2003). *Ergebnisse aus IGLU: Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. New York: Waxmann.

Mullis, I.V.S., Martin, M.O., & Gonzalez, E.J. (2004). *PIRLS international achievement in the processes of reading comprehension: Results from PIRLS 2001 in 35 countries*. Chestnut Hill, MA: Boston College.

for reading. To support the creation of two reading process scales, the total assessment time required needed to increase, and the booklet design expanded to include additional test booklets.

The decision to report reading achievement scale scores by process as well as by purpose, in combination with the desire to include a range of texts within each reading purpose, made it necessary to increase PIRLS 2006 to include five literary and five informational test blocks. Each of the 10 test blocks included a reading passage and its accompanying questions. As shown in Exhibit 2.3, half of the test blocks were devoted to measuring literary purposes (labeled L1-L5), and the other half were focused on acquiring and using information (labeled I1-I5). Since four of the PIRLS 2001 test blocks were kept secure and carried forward for measuring for trends in 2006, development efforts in 2006 focused on the six remaining blocks.

Exhibit 2.3 PIRLS 2006 Student Booklet Design

Literary Block Number	Literary Title	Informational Block Number	Informational Title
L1	Lump of Clay	I1	Antarctica
L2	Flowers	I2	Leonardo
L3	To be developed	I3	To be developed
L4	To be developed	I4	To be developed
L5	To be developed	I5	To be developed

2.5 Finding and Selecting Passages

Development of the PIRLS 2006 reading assessment involved selecting passages from existing sources representative of the types of materials likely to be read by children in the fourth grade, writing items based on these texts, and devising scoring guides for constructed-response items. These new materials were to reflect the broad approaches established for PIRLS 2001, while refreshing and expanding the range of texts and devising items that brought out the qualities of each passage.

The task of selecting passages for an international assessment is a demanding one. In this case, it was desirable that the new texts have features in line with the framework and maintaining a recognizable continuity with the secure passages from PIRLS 2001. In undertaking all stages of this task, collaboration among the participating countries was a central part of the work.

Based on the need for 6 new text blocks of passages and items, it was decided to develop 12 such blocks for the field test. A call for passages was sent out to all NRCs, with a request for submission of both literary and informational reading passages to be used as the foundation for the development of test items. Research coordinators were asked to submit passages with the following characteristics:

- Suitable for fourth-grade students in content, interest, and reading ability;
- Well written in terms of depth and complexity to allow questioning across the processes and strategies defined in the PIRLS 2006 framework; and
- Sensitive to cultural groups to avoid specific cultural references, wherever possible.

The text of the passages, written in or translated into English, had to be continuous and not exceed 1,200 words. Examples of literary text include short stories, narrative extracts, traditional tales, fables, myths, and play scripts. Informational texts include textbook or expository passages, biographies, and persuasive writing and could include charts, tables, or diagrams.

To begin with, the TIMSS & PIRLS International Study Center received over 50 reading passages from NRCs in the following countries: Canada, the Czech Republic, England, France, Germany, Hong Kong SAR, Iran, Italy, New Zealand, and Singapore. Passages were circulated and reviewed at the first meeting of the coordinators. Of the passages reviewed, they chose four literary and six informational texts to be revised and edited for the field test. After examining themes and content of the selected passages, the coordinators agreed to continue to find and submit additional suitable passages that varied in style and structure.

The PIRLS Reading Development Group convened for the first time in January 2004 to review passages selected by the NRCs, as well as those additional passages submitted afterwards. Reading Development Group members made suggestions for minor text revisions for consistency of language and drafted a preliminary list of possible questions for each passage. In total, the Reading Development Group recommended eight literary and eight informational passages for the next stage of development, an item-writing workshop at the second meeting of NRCs. The passages differed by content, style, and length. The

literary passages ranged in length from 797 words to 1,127 words and included a mixture of traditional and contemporary stories with an array of characters and story plots. Informational passages ranged in length from 693 to 985 words and represented a wide range of topics and informational text structures. For each of the 16 passages presented, the Item Development Task Force constructed a text map highlighting the passage's central ideas and key features.

2.6 Developing Items and Scoring Guides

In February 2004, the TIMSS & PIRLS International Study Center convened a meeting of the NRCs to review the set of 16 passages recommended by the Reading Development Group and to write items and scoring guides for those considered most suitable. From the 16, the coordinators selected a subset of 13 passages—seven literary and six informational—for which they would create items and scoring guides for constructed-response items and recommended that a seventh informational passage be identified before the field test.

The workshop began with basic training in developing reading items. As a basis for the training, the TIMSS & PIRLS International Study Center provided NRCs with *Item-Writing Guidelines for the PIRLS 2006 Field Test* (2004). The guidelines were reviewed and discussed, including general issues for writing items and scoring guides, a system for documenting and classifying items for review, and procedures for reviewing items and scoring guides once they were written. The following is a summary of item-writing guidelines for each passage:

- Write items totaling at least 18-20 score points (approximately 12–13 items) per passage.
- Write questions that match the purpose of the passage as classified for PIRLS 2006, paying close attention to writing questions that cover the range of the four PIRLS comprehension processes.
- Write questions relevant to the central ideas in the passage, making sure that answering a question correctly depends on having read the passage.
- Each item should be independent of the other items (no item should provide “clues” to the correct response for another item).
- For each question, consider the timing, grade appropriateness, difficulty level, potential sources of bias, and ease of translation.

- For multiple-choice questions, ask direct questions, making sure there is one and only one correct answer and provide plausible distracters.
- Develop a unique, tailored scoring guide for each constructed-response item. Write a full-credit answer to each question in terms of language, knowledge, and skills of a typical fourth-grade student.

Guidelines were reviewed for constructing unique scoring guides for 1-, 2-, and 3-point constructed-response items. For each item, scoring criteria were to be as specific as possible in order to standardize scoring decisions across countries, as well as provide for a range of responses within each score level. The guidelines emphasized the features required within each scoring guide for each score level:

- A general statement about the nature of comprehension, which is characteristic of responses at that level,
- Specific content of students' responses that may be considered evidence of an appropriate inference, and
- Examples of the various types of plausible student responses.

The NRCs divided into eight groups with three to six people per group. Each of the groups was assigned at least one literary and one informational passage. At least two groups worked on each passage, in order to maximize the number and variety of items drafted for each passage. Each group reviewed their sets of items and made final revisions before supplying an electronic file of all items to the TIMSS & PIRLS International Study Center. By the end of the meeting, coordinators had drafted a total of 277 items, as well as scoring guides for all constructed-response items.

The TIMSS & PIRLS International Study Center combined and organized the draft items by passage, keeping items with similar topics or themes together and distributed the items in March 2004 to the Item Development Task Force for review. The review was based on a simple rating system of 1 to 3, ranging from exceptional or requiring minimal revision to requiring extensive modifications. Task Force members were invited to comment on individual items, as well as provide a general overview of the set of items for a particular passage. The Task Force met for 3 days in June 2004 to review the draft item pool and make suggestions for revisions and recommendations of items that should be retained for the field test. In addition to the item review, the Task Force drafted items for a new informational passage that had been identified and approved by

NRCs after their second meeting. Throughout the months of June and July, the Task Force continued to refine the field-test items for review by the Reading Development Group.

In August 2004, the Reading Development Group convened to evaluate the 14 passages and accompanying items developed to date, and recommend the 12 most suitable blocks for the field test (see Exhibit 2.4 for a list of titles). Reading Development Group members made minor suggestions for edits to the passages and questions, primarily to refine the intent of questions or how individual constructed-response items should be scored. Another focus of the Reading Development Group meeting was to ensure proper balance in terms of the processes measured, item types, and total number of points across the passages.

Exhibit 2.4 PIRLS 2006 Field Test Passages

Literary Title	Informational Title
Shiny Straw	Spacewalking
The Fox and the Rooster	Sharks
Fly Eagle	Searching for Food
Unbelievable Night	Chocolate Then and Now
Growing Money	Day Hiking
Dolphin Rescue	Ice Age Cave

There was a NRC meeting in November 2004 before conducting the field test to finalize and approve the field-test materials for production and administration. The NRCs generally endorsed the versions of the assessment blocks as revised by the Reading Development Group, with some proposed alterations to two of the informational passages. Other minor changes were made to items in the literary blocks to improve clarity of the questions and options. Following the meeting, the TIMSS & PIRLS International Study Center implemented the suggested changes and provided the final international version of the PIRLS 2006 field-test booklets to the NRCs on December 1, 2004.

2.7 Conducting the PIRLS 2006 Field Test

In preparation for the assessment data collection in 2006, PIRLS conducted a full-scale field test between March and April 2005. The field-test data provided a good basis for evaluating the measurement properties of the new assessment

blocks developed in the 2006 round. The 12 assessment blocks developed for the field test were paired into six booklets, each booklet containing one literary and one informational block. To ensure the target sample size for a minimum of 200 student responses per item, participating countries sampled between 25 and 40 schools, depending upon class sizes. In total, nearly 50,000 students from more than 1,200 schools in 42 countries participated in the field test.

2.8 Piloting Items for Scoring Guides

The PIRLS 2006 constructed-response items permit a wide range of student responses. Young students express their understanding in various ways, and, in some cases, the justification for an answer can take different forms. Because scoring these varying responses must be carried out consistently, extensive training in applying the scoring guides required collecting student responses to use as training materials. These student responses also were helpful for refining the scoring guides.

To refine the scoring guides for constructed-response items and prepare scoring training materials, seven English-speaking countries agreed to administer field-test booklets to a small selection of classrooms. In total, approximately 200 student responses to each of the constructed-response items were collected from the Canadian provinces of Ontario and Quebec, England, New Zealand, Scotland, Singapore, South Africa, and the United States.

The PIRLS Item Development Task Force met at the TIMSS & PIRLS International Study Center in February 2005 to evaluate the student responses and make adjustments to the scoring criteria to accommodate appropriate responses not accounted for in the existing scoring guides. The Task Force scored the student responses according to the guides and grouped the responses to each item by score category. Next, the Task Force reassessed the sets of scored responses and reconciled scores for responses not easily categorized. To create sets of student responses for training scorers, the Task Force selected from among the scored responses from the item pilot.

2.9 Scoring Training for Constructed-response Items from the PIRLS 2006 Field Test

In March 2005, the TIMSS & PIRLS International Study Center held a meeting for the NRCs and their scoring managers who would implement the constructed-response scoring in each participating country. The majority of the meeting

consisted of a 4-day training session in the application of the scoring guides for constructed-response items for the field test. For each item, a set of training materials was provided. The training materials included 8–10 anchor papers and a set of 8–10 practice papers for each of the 76 constructed-response items, arranged by assessment block. Sets of anchor and practice papers contained student responses collected from the item pilot.

Following the review of the text for each scoring guide, the participants were provided with a set of anchor papers comprised of at least three example student responses for each of the score-point categories for an item. Rationales for the score assigned to each anchor paper were included for the set and discussed during the presentation of each example. Upon completion of the review of anchor papers for an item, the NRCs and scorers read unscored student responses in the practice paper sets and participated in an open discussion of rationales for scoring the practice items.

2.10 Selecting Final Reading Passages for the PIRLS 2006 Data Collection

The Reading Development Group met in July 2005 to study the results of the PIRLS 2006 field test and recommend three literary and three informational blocks for inclusion in the main data collection. Criteria for recommendations for the assessment included desirable overall passage statistics and individual item statistics in addition to well-suited blocks representative of the reading experiences of fourth graders internationally. Item statistics were used to evaluate the effectiveness of the items and identify items requiring revisions. The group was generally pleased with the measurement characteristics of the items and proposed 6 of the 10 blocks be presented to the NRCs for inclusion in PIRLS 2006 and two that would be used in the PIRLS framework as examples of the PIRLS assessment.

The NRCs reviewed the item statistics from the field test at their fifth meeting in August 2005. An extensive discussion of the field-test results and the qualities of the secured PIRLS 2001 trend blocks, which also would appear in the 2006 assessment, led NRCs to adopt the recommendations of the Reading Development Group, with the provision that one literary block recommended for the framework replace another block recommended for the assessment. Based on item statistics, a number of individual items were identified by the coordinators and modified to improve clarity and accuracy of student responses.

Following the review, TIMSS & PIRLS International Study Center staff implemented edits to the blocks. The new assessment blocks developed for PIRLS 2006 were combined with the secure blocks from the 2001 assessment, providing an overall assessment that would allow the calculation of trends over 5 years, as well as containing new material. Finalized assessment materials were made available to the NRCs on August 15, 2005, in preparation for the main data collection, which began in Southern Hemisphere countries in October 2005. Exhibit 2.5 lists the PIRLS 2006 passage titles by block.

Exhibit 2.5 PIRLS 2006 Student Booklet Design

Literary Block Number	Literary Title	Informational Block Number	Informational Title
L1	Lump of Clay (2001)	I1	Antarctica (2001)
L2	Flowers (2001)	I2	Leonardo (2001)
L3	Shiny Straw (2006)	I3	Day Hiking (2006)
L4	Fly Eagle (2006)	I4	Sharks (2006)
L5	Unbelievable Night (2006)	I5	Searching for Food (2006)

The PIRLS 2006 assessment included 126 items across the 10 assessment blocks, comprising a total of 167 score points. The numbers of multiple-choice and constructed-response items by reading purpose are presented in Exhibit 2.6. The two question formats—constructed response and multiple choice—were evenly represented in the total number of items, with 64 multiple-choice items and 62 constructed-response items in the assessment. The total number of items and score points were distributed equally between the two purposes for reading.

Exhibit 2.6 PIRLS 2006 Assessment Item Specifications

	Number of Multiple-choice Items	Number of Constructed-response Items			Total Number of Items	Total Number of Score Points
		1 pt.	2 pts.	3 pts.		
Literary	34	13	13	4	64	85
Informational	30	15	14	3	62	82
Total	64	28	27	7	126	167

Exhibit 2.7 presents the portion of the assessment devoted to each of the four processes of reading comprehension. The distribution of actual score points across the processes approximates the distribution established in the PIRLS 2006 framework. Equal proportions among the first two and last two processes support the reporting of separate scales for two processes of comprehension: retrieval and straightforward inferencing and interpreting, integrating, and evaluating.³

Exhibit 2.7 Distribution of Score Points Across Reading Processes

PIRLS 2006 Processes of Reading Comprehension	Number of Score Points	Percentage of Total Score Points
Focus on and retrieve explicitly stated information	36	22
Make straightforward inferences	47	28
Interpret and integrate ideas and information	61	37
Examine and evaluate content, language, and textual elements	23	14
Total	167	100

2.11 Finalizing the PIRLS 2006 Scoring Guides for Constructed-response Items

In October 2005, the PIRLS Item Development Task Force met to review and revise the constructed-response scoring guides and sets of training materials in response to changes made to items after the field test. Most constructed-response items required minor changes. Only two constructed-response items and their scoring guides, each from a different assessment block, were modified significantly. After the meeting, a small number of classes from England and Scotland were administered a pilot test consisting of the two assessment blocks, and the student responses were used to rewrite the scoring guides and provide examples for the scoring training materials for the main data collection.

In addition, the Task Force reviewed all the examples and training materials for coherence and consistency, in light of responses from the field test and to ensure that the characteristic patterns of student response were covered by the guides.

These final versions of the scoring guides and training materials from the PIRLS 2006 field test were combined with those from the 2001 passages

³ Retrieval and straightforward inferencing will combine items from the focus on and retrieve explicitly stated information and make straightforward inferences comprehension processes. Similarly, interpreting, integrating, and evaluating will be based on items from the interpret and integrate ideas and information and examine and evaluate content, language, and textual elements processes.

that now were included in the 2006 assessment. As in the field test, training materials, which included 8–10 anchor papers and 8–10 practice papers for each of the 62 constructed-response items, were arranged by assessment block. All scoring guides and training materials then were introduced to the NRCs and their scoring managers in two intensive training sessions in November 2005 for the Southern Hemisphere and March 2006 for the Northern Hemisphere. Discussion of the student responses in the training materials allowed the participants in the training sessions to become confident with the distinctions between the various levels of scoring categories and to pass this knowledge on to their scoring teams.

References

- Campbell, J.R., Kelly, D.L., Mullis, I.V.S., Martin, M.O., & Sainsbury, M. (2001). *Framework and specifications for PIRLS assessment 2001* (2nd ed.), Chestnut Hill, MA: Boston College.
- Mullis, I.V.S., Kennedy, A.M., Martin, M.O., & Sainsbury, M. (2006). *PIRLS 2006 assessment framework and specifications* (2nd ed.), Chestnut Hill, MA: Boston College.
- TIMSS & PIRLS International Study Center. (2004). *Item-writing guidelines for the PIRLS 2006 field test*. Chestnut Hill, MA: Boston College.