



# Overview of PIRLS

---

Dana L. Kelly

## 1.1 Background

IEA has been conducting cross-national studies of educational achievement for more than 40 years – including periodic assessments of children’s reading literacy. In 1973, reading was one of the subjects in IEA’s six-subject study, which was conducted in 15 countries (Thorndike, 1973; Walker, 1976). In 1991, IEA’s Reading Literacy Study was conducted in 32 educational systems (Elley, 1992; 1994). Most recently, PIRLS (Progress in International Reading Literacy Study) was established as the IEA’s latest study to monitor progress in children’s reading literacy into the future (Mullis, Martin, Gonzalez, & Kennedy, 2003).

In 1998, the IEA General Assembly formally agreed that PIRLS would be part of the IEA’s regular cycle of assessments, which also includes mathematics and science. At that point, some basic principles were established.

- PIRLS would begin in 2001 with an assessment of children in fourth grade.
- PIRLS would focus on reading literacy achievement, as well as home and school contexts for learning to read.
- Reading literacy would be measured through a comprehensive assessment based on authentic reading materials requiring students to engage in a range of reading processes.
- The reading test would be designed so that future assessments could measure trends in achievement.

- Questionnaires would be administered to the tested students, their current reading teachers, and their school principals – to collect contextual data with which to interpret achievement.

In 1999, planning for the study began with a meeting among representatives from the IEA Secretariat, the International Study Center (ISC) at Boston College, Statistics Canada, and the National Foundation for Educational Research in England and Wales. At this meeting, it was established that – in addition to incorporating the General Assembly’s basic principles – PIRLS would try to collect data from children’s parents about literacy activities in the home, and also collect data about early reading instruction in schools to provide additional information on reading instruction (beyond what the current-year teachers would provide). These basic goals were supported by the Reading Development Group (RDG) and representatives from the participating countries (the National Research Coordinators).

The development of PIRLS spanned two years, beginning in 1999 and continuing until early 2001, when the final reading test and questionnaires were approved by the participating countries. As part of development, 30 countries conducted a field test of the test and questionnaires. Ultimately, 35 countries participated in the main data collection.

## 1.2 Participating Countries

Thirty-five countries joined together to conduct the first PIRLS assessment in 2001:

Argentina	Latvia
Belize	Lithuania
Bulgaria	Macedonia
Canada (Ontario, Quebec)	Moldova
Colombia	Morocco
Cyprus	Netherlands
Czech Republic	New Zealand
England	Norway
France	Romania
Germany	Russian Federation
Greece	Scotland
Hong Kong	Singapore
Hungary	Slovak Republic
Iceland	Slovenia
Iran	Sweden
Israel	Turkey
Italy	United States
Kuwait	

## 1.3 Student Population Assessed

In 2001, PIRLS assessed the reading literacy of children in “the upper of the two grades with the most 9-year-olds at the time of testing” (PIRLS, 1999). This corresponds to the fourth grade in most countries. This population was chosen because it represents an important transition point in children’s development as readers. In most countries, by the end of fourth grade, children are expected to have learned how to read, and are now reading to learn. This grade is also assessed in the IEA’s Trends in International Mathematics and Science Study (TIMSS), to

provide countries participating in both studies with achievement and background data for three subjects at the same grade level.

In each country, representative samples of students were selected using a two-stage sampling design. In the first stage, at least 150 schools were selected using probability-proportional-to-size sampling. Countries could incorporate in their sampling design important reporting variables (for example, urbanicity or school type) as stratification variables. At the second stage, one or two fourth-grade classes were randomly sampled in each school. This resulted in a sample size of at least 3,750 students in each country. Some countries opted to include more schools and classes, enabling additional analyses, which resulted in larger sample sizes.

#### 1.4 Assessment Dates

PIRLS was administered near the end of the school year in each country. In countries in the Northern Hemisphere (where the school year typically ends in May or June) the assessment was conducted in April, May, or June 2001. In the Southern Hemisphere, the school year typically ends in November or December; so in these countries, the assessment was conducted in October or November 2001.

#### 1.5 Study Management and Organization

PIRLS is directed in the United States by Ina V.S. Mullis and Michael O. Martin, at the International Study Center at Boston College; they also direct the IEA's TIMSS. The PIRLS International Study Center was responsible for the design, development, and implementation of the study – including developing the instruments and survey procedures; ensuring quality in data collection; and analyzing and reporting the study results. The International Study Center worked closely with the organizations responsible for particular aspects of the study, the PIRLS advisory committees, and representatives of the participating countries.

Each country appointed a National Research Coordinator (NRC) who, together with staff members at the PIRLS national center, was responsible for all aspects of the study within that country. The PIRLS ISC organized meetings of the NRCs several times a year to review study materials and procedures, and to receive training in scoring constructed-response items, and in entering the data using the prescribed software.

The IEA Secretariat provided guidance in all aspects of the study, and was responsible for managing the ambitious translation verification effort conducted for the field test and main assessment. Statistics Canada was responsible for all aspects of sampling – including working with countries to ensure that the international procedures are followed; adapting the international

design to national conditions; documenting the national samples; and computing sampling weights.

The National Foundation for Educational Research in England and Wales had major responsibility for developing the reading test – including collecting reading passages; developing items and scoring guides; and conducting scoring training. The IEA Data Processing Center was responsible for processing and verifying the data from the 35 countries and for constructing the international database. Educational Testing Service provided software and support for scaling the achievement data.

The study directors and representatives from the International Study Center, IEA, Statistics Canada, the National Foundation for Educational Research in England and Wales, and Educational Testing Service met periodically to review the study's progress, procedures, and schedule.

The PIRLS Reading Development Group (see Appendix A) contributed their invaluable expertise to the framework and reading test. Committee members reviewed various drafts of the framework and assessment blocks, and reviewed and endorsed the final reading test. The PIRLS Questionnaire Development Group (see Appendix A) – comprising representatives from six countries – helped develop the PIRLS questionnaires (including writing items and reviewing drafts of all questionnaires).

## 1.6 Overview of Assessment Framework

At the heart of the PIRLS assessment is the definition of reading literacy established by the Reading Development Group, and refined by National Research Coordinators. The PIRLS definition of reading literacy builds on the definition used in the IEA 1991 study, but elaborates on that definition by making specific reference to reading by children. PIRLS defines reading literacy 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, and for enjoyment (Campbell, Kelly, Mullis, Martin, & Sainsbury, 2001).

Growing out of this definition are the three aspects of reading literacy assessed by PIRLS:

- Processes of comprehension
- Purposes for reading and
- Reading behaviors and attitudes.

Processes of comprehension and purposes for reading are the foundation of the written assessment of reading comprehension. The purposes for reading and processes of comprehension, as well as the percentages of the assessment devoted to each, are shown in Exhibit 1.1. Each process is assessed with each purpose for reading. Reading behaviors and attitudes are assessed through a questionnaire completed by the students.

**Exhibit 1.1: Percentages of Reading Assessment Devoted to Reading Purposes and Processes<sup>1</sup>**

		Purposes for Reading	
		Literary Experience	Acquire and Use Information
Processes of Comprehension	Focus on and Retrieve Explicitly Stated Information	9%	13%
	Make Straightforward Inferences	14%	9%
	Interpret and Integrate Ideas and Information	20%	20%
	Examine and Evaluate Content, Language, and Textual Elements	6%	8%

<sup>1</sup> Because numbers are rounded to the nearest whole number, some totals may appear inconsistent.

## 1.7 PIRLS Reading Assessment

PIRLS has ambitious goals for covering the domain of reading literacy. The Reading Development Group felt that at least eight passages and items (four for each reading purpose) were needed to provide a valid and reliable measure of reading achievement. Since it would not be possible to administer the entire test to any one child, PIRLS used a matrix sampling technique to distribute the assessment material among students, yet retain linkages necessary for scaling the achievement data.

**Exhibit 1.2: PIRLS Assessment Design**

	Booklet 1	Booklet 2	Booklet 3	Booklet 4	Booklet 5	Booklet 6	Booklet 7	Booklet 8	Booklet 9	Booklet R (Reader)
Assessment Block	L1	L2	L3	I1	I2	I3	L1	I2	I3	L4
	L2	L3	I1	I2	I3	L1	I1	L2	L3	I4

L = Reading for Literary Experience; I = Acquire and Use Information

### 1.7.1 Assessment Design

The material was divided into 40-minute “blocks,” each comprising a passage (a story or article) and items representing at least 15 score points. There are eight such blocks, four for each reading purpose. Blocks containing literary passages are labeled L1 through L4, and those containing informational passages, I1 through I4. The eight assessment blocks are distributed across ten test booklets, and each student completed one booklet in an 80-minute testing session. Each booklet contains two blocks, and most blocks appeared in three booklets. One of the ten booklets is the PIRLS Reader, a color booklet containing two reading passages; the test items are located in a separate booklet. The two blocks comprising the Reader appear only in that booklet. The distribution of blocks across booklets “links” the booklets to enable the achievement data to be scaled using item response theory methods.

The design for the assessment booklets is presented in Exhibit 1.2, which shows that each booklet has two blocks – two literary, two informational, or one of each. It also shows that three of the literary and three of the informational blocks appear three times

**Exhibit 1.3:** Distribution of Items by Type and Reading Purpose

	Multiple-Choice Items	Constructed-Response Items			Total Number of Items	Total Score Points
		1-point	2-point	3-point		
Literary	25	14	9	3	51	66
Informational	21	10	12	4	47	67
Total	46	24	21	7	98	133

across test booklets 1 through 9, and that the fourth literacy and informational blocks appear only in the Reader.

### 1.7.2 Passages

The reading passages form the foundation of the reading literacy test. In accordance with the framework, four of the assessment blocks contain literary texts and four contain informational texts, and the passages are authentic texts drawn from children’s storybooks and informational sources. Submitted and reviewed by the PIRLS countries, the passages represent a range of types of literary and informational texts. The literary passages include realistic stories and traditional tales; while the informational texts include chronological and non-chronological articles, a biographical article, and an informational leaflet.

### 1.7.3 Items and Scoring Guides

Two item formats were used to assess children’s reading literacy – multiple-choice and constructed-response. Each type of item was used to assess both reading purposes and all four reading processes. Multiple-choice items provided students with four possible answers, one of which was correct. Each multiple-choice item was worth one point. Constructed-response items required students to construct their answers rather than select from among

possible answers. These items were worth one, two, or three points – depending on the depth of understanding or extent of textual support the item required.

Each block of assessment material contained from 11 to 14 items that together represent at least 15 score points. Altogether, the PIRLS reading test includes 98 items representing 133 score points – enough to estimate achievement reliably. Exhibit 1.3 shows the distribution of items by type and reading purpose.

Scoring guides for constructed-response items were developed together with the items. Each scoring guide is unique to that item. It describes the essential features of appropriate and complete responses – including the kind of evidence of understanding required and example student responses to help scorers determine the score for a particular response. Actual student responses were used to develop the guides, and illustrate the kinds of responses garnering different points.

### 1.7.4 Releasing Assessment Material to the Public

The PIRLS test design provides for the release of half of the assessment material into the public domain after data collection, including the entire PIRLS Reader.

The remaining half will be kept secure and included in future PIRLS assessments so trends in achievement can be measured. After data collection in 2001, one literary and one informational block (as well as the blocks in the Reader) were released. Effectively, four blocks (or half of the assessment) were available to the public.

## **1.8 PIRLS Background Questionnaires**

By gathering information about children's experiences together with reading achievement on the PIRLS test, it is possible to identify the factors or combinations of factors that relate to high reading literacy. An important part of the PIRLS design is a set of questionnaires targeting factors related to reading literacy. PIRLS administered four questionnaires: to the tested students, to their parents, to their reading teachers, and to their school principals.

### **1.8.2 Student Questionnaire**

Each student taking the PIRLS reading assessment completes the student questionnaire. The questionnaire asks about aspects of students' home and school experiences – including instructional experiences and reading for homework, self-perceptions and attitudes towards reading, out-of-school reading habits, computer use, home literacy resources, and basic demographic information.

### **1.8.3 Learning to Read (Home) Survey**

The learning to read survey is completed by the parents or primary caregivers of each student taking the PIRLS reading assessment. It addresses child/parent literacy

interactions, home literacy resources, parents' reading habits and attitudes, home/school connections, and basic demographic and socioeconomic indicators.

### **1.8.4 Teacher Questionnaire**

The reading teacher of each fourth-grade class sampled for PIRLS completes a questionnaire designed to gather information about classroom contexts for developing reading literacy. This questionnaire asks teachers about characteristics of the class tested (such as size, reading levels of the students, and the language abilities of the students). It also asks about instructional time, materials and activities for teaching reading and promoting the development of their students' reading literacy, and the grouping of students for reading instruction. Questions about classroom resources, assessment practices, and home/school connections also are included. The questionnaire also asks teachers for their views on opportunities for professional development and collaboration with other teachers, and for information about their education and training.

### **1.8.5 School Questionnaire**

The principal of each school sampled for PIRLS responds to the school questionnaire. It asks school principals about enrollment and school characteristics (such as where the school is located, resources available in the surrounding area, and indicators of the socioeconomic background of the student body), characteristics of reading education in the school, instructional time, school resources (such as the availability of instructional materials and staff), home/school connections, and the school climate.

## 1.9 Translation and Verification of Instruments

The PIRLS reading tests and questionnaires were prepared in English, then translated into 31 other languages for use in the 35 participating countries. Countries were responsible for translating the instruments into their local language (or languages) following internationally prescribed procedures. To ensure standardization of instruments across countries, PIRLS undertook an extensive verification process, whereby each country's data collection instruments were independently reviewed and verified by an external translation company engaged by the IEA. The verifiers' reviews of the translated documents were used to improve the translations. Instruments were verified twice, once before the field test and once before the main data collection. In addition to the external review, the International Study Center also reviewed the countries' instruments against the verifiers' comments to ensure that all necessary corrections were made. Finally, statistical analyses of item data were conducted to check for evidence of differences in student performance across countries that could indicate a translation problem.

## 1.10 Data Collection

Each country was responsible for carrying out all aspects of the data collection, using standardized procedures developed for the study. Manuals provided explicit instructions to the NRCs and their staff members on all aspects of the data collection – from

contacting sampled schools to packing and shipping materials to the IEA Data Processing Center for processing and verification. Manuals were also prepared for test administrators and for individuals in the sampled schools who work with the national centers to arrange for the data collection within the schools. These manuals addressed all aspects of the assessment administration within schools (including test security, distribution of booklets, timing and conduct of the testing session, and returning materials to the national center).

The PIRLS International Study Center placed great emphasis on monitoring the quality of the PIRLS data collection. In particular, the Study Center implemented an international program of site visits, whereby international quality control monitors visited a sample of 15 schools in each country and observed the test administration. In addition to the international program, NRCs were also expected to organize an independent national quality control program based upon the international model. The latter program required national Quality Control Observers to document data collection activities in their country. The national Quality Control Observers visited a random sample of 10 percent of the schools (in addition to those visited by the international Quality Control Monitors), and monitored the testing sessions – recording their observations for later analysis.



### 1.11 Scoring the Constructed-Response Items

Because almost two-thirds of the score points came from constructed-response items, PIRLS needed to develop procedures for reliably evaluating student responses within and across countries. The International Study Center prepared detailed guides containing the PIRLS scoring rubrics, and explanations of how to implement them – together with example student responses for the various rubric categories. These guides, along with training packets containing extensive examples of student responses for practice in applying the rubrics, were used as a basis for intensive training of national representatives in scoring the constructed-response items.

To gather and document empirical information about the within-country agreement among scorers, PIRLS arranged to have a sample of 200 students' responses to each item in each country scored independently by two readers. Scoring reliability within countries was high – the percentage of exact agreement, on average, across countries, was more than 90 percent. PIRLS also conducted a study of scoring reliability across countries, asking countries with scorers proficient in English to score a reference set of student responses chosen from students in English-speaking countries. This study revealed a high level of agreement between scorers also (85% on average).

### 1.12 Data Processing

To ensure the availability of comparable, high-quality data for analysis, PIRLS took rigorous quality control steps to create the international database. PIRLS prepared manuals and software for countries to use in creating and checking their data files, so that the information would be in a standardized international format before being forwarded to the IEA Data Processing Center (DPC) in Hamburg for creation of the international database. Upon arrival at the DPC, the data underwent an exhaustive cleaning process involving several iterative steps and procedures designed to identify, document, and correct deviations from the international instruments, file structures, and coding schemes. The process also emphasized consistency of information within national data sets, and appropriate linking among the student, parent, teacher, and school data files.

Throughout the process, the data were checked and double-checked by the IEA Data Processing Center, the International Study Center, and the national centers. The national centers were contacted regularly and given multiple opportunities to review the data for their countries. In conjunction with the IEA Data Processing Center, the International Study Center reviewed item statistics for each cognitive item in each country to identify poorly performing items. In general, the items exhibited very good psychometric properties in all countries.

### 1.13 IRT Scaling

The general approach to reporting the PIRLS achievement data was based primarily on item response theory (IRT) scaling methods. Student reading achievement was summarized using a family of IRT models (2-parameter, 3-parameter, and generalized partial credit models). The IRT methodology was preferred for developing comparable estimates of performance for all students, since students responded to different passages and items depending upon which of the test booklets they received (Booklet 1 through 9 or the PIRLS Reader). This methodology produces a score by averaging the responses of each student to the items that he or she took in a way that takes into account the difficulty and discriminating power of each item. The approach followed in PIRLS uses information from the background questionnaires to provide improved estimates of student performance (a process known as conditioning) and multiple imputation to generate student scores (or “plausible values”) for analysis and reporting.

The IRT analysis provides a common scale on which performance can be compared across countries. In addition to providing a basis for estimating mean achievement, scale scores permit estimates of how students within countries vary and provide information on percentiles of performance. Treating all participating countries equally, the PIRLS scale average across countries was set to 500 and the standard deviation to 100. Since the countries varied in size, each country was weighted to contribute equally to the mean and standard deviation of the

scale. The average and standard deviation of the scale scores are arbitrary and do not affect scale interpretation.

In the PIRLS analysis, achievement scales were produced for each of the two reading purposes, reading for literary experience and reading for information, as well as for reading overall.

### 1.14 Data Analysis and Reporting

The PIRLS 2001 International Report (Mullis, Martin, Gonzalez & Kennedy, 2003) summarizes fourth-grade students’ student reading achievement in each country. This report presents average student achievement in reading overall as well as in reading for literary experience and reading to acquire and use information, together with standard errors and tests of significance as appropriate. Average achievement is reported separately for girls and boys.

To provide additional information about reading achievement among high- and low-achieving students, PIRLS reported the percentage of students in each country performing at each of four international benchmarks of student achievement – corresponding to the 90th, 75th, 50th, and 25th percentiles of the international distribution of reading achievement. To enhance this reporting approach, PIRLS conducted a scale anchoring analysis to describe student performance at the international benchmarks in terms of the kind of reading students performing at each benchmark can do, and the level of comprehension they exhibit. Complementing this approach fur-

ther, the PIRLS international report presents examples of questions from both literary and informational passages that anchor at each of the benchmarks (providing another perspective on the reading demands of the benchmarks), and also displays student performance in each country on the example questions.

PIRLS 2001 collected a wide array of information about the home and school context in which students learned to read (from parents, students, teachers, and school principals). The PIRLS international report summarized much of this information, combining data into composite indices showing an association with achievement where appropriate. In particular, student reading achievement is described in relation to literacy-related activities in the home, the school curriculum and organization for teaching reading, teachers and reading instruction, school contexts, and students' reading attitudes, self-concepts, and out-of-school activities.

Additional information about the countries participating in PIRLS 2001 may be found in the *PIRLS 2001 Encyclopedia* (Mullis, Martin, Kennedy, Flaherty, 2002), a volume providing general information on the cultural, societal, and economic situation in each country, as well as a more focused perspective on the structure and organization (of their respective educational systems as it pertains specifically to the promotion of reading literacy). Consisting of a chapter from each country, the *PIRLS 2001 Encyclopedia* describes primary/elementary schooling as it pertains to reading within each educational system – including teacher education and training,

reading curricula, classroom organization and instruction, and assessment practices. As such, it is an extremely valuable companion publication to the international report providing insights and detailed information about the policies, practices, and resources within each country.

### 1.15 The Trends in IEA's Reading Literacy Study

While PIRLS 2001 is the first in a cycle of assessments designed to measure trends in reading achievement, some countries also measured achievement trends from 1991 to 2001. Countries that participated in the IEA's 1991 Reading Literacy Study were eligible to administer the 1991 reading test and student questionnaire to a sample of students in 2001 so that they could obtain information about how their children's reading literacy today compares with that of ten years ago. The following countries participated in the trend study:

Greece	Singapore
Hungary	Slovenia
Iceland	Sweden
Italy	United States
New Zealand	

Countries sampled every other PIRLS school for the trend study, resulting in a sample size of at least 75 schools. In each school, one target-grade classroom was sampled and administered the 1991 test and student questionnaire. For some countries, the 1991 target grade and the PIRLS target grade were not the same. Statistics Canada worked with these countries to tailor the design so

as to achieve a representative sample of students. The IRT scaling methodology used with PIRLS 2001 also was applied in the trends in reading literacy study. The results of the trend study are presented in Martin, Mullis, Gonzalez, and Kennedy (2003).

## 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.
- Elley, W.B. (1992). *How in the world do students read?* The Hague, Netherlands: IEA.
- Elley, W.B. (Ed.). (1994). *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems*. Oxford, England: Elsevier Science Ltd.
- Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., & Kennedy, A.M. (2003). *Trends in children's reading literacy achievement 1991-2001: IEA's repeat in nine countries of the 1991 Reading Literacy Study*. Chestnut Hill, MA: Boston College.
- Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Kennedy, A.M. (2003). *PIRLS 2001 International report: IEA's study of reading literacy achievement in primary schools in 35 countries*. Chestnut Hill, MA: Boston College.
- Mullis, I.V.S., Martin, M.O., Kennedy, A.M., & Flaherty, C.L. (2002). *PIRLS 2001 Encyclopedia: A reference guide to reading education in the countries participating in IEA's Progress in International Reading Literacy Study (PIRLS)*. Chestnut Hill, MA: Boston College.
- PIRLS (1999). *School Sampling Manual – Version 2* (PIRLS Ref. No. 99-0019). Prepared by Pierre Foy & Marc Joncas, Statistics Canada. Chestnut Hill, MA: Boston College.
- Thorndike, R.L. (1973). Reading comprehension in fifteen countries: An empirical study. *International studies in evaluation: Vol. 3*. Stockholm: Almqvist & Wiksell.
- Walker, D.A. (1976). *The IEA six subject survey: An empirical study of education in twenty-one countries*. New York: John Wiley & Sons Inc.