

PIRLS

Appendix A

A

Appendix A

Overview of PIRLS Procedures

History

With the Progress in International Reading Literacy Study (PIRLS), the International Association for the Evaluation of Educational Achievement (IEA) continues its mission to conduct comparative studies of student achievement in school subjects to inform educational policies and practices in countries around the world. Since its inception in 1959, the IEA has conducted studies on a wide range of topics and subjects including mathematics, science, language, civics, and reading. By providing a cross-national perspective on educational systems and on organizational and instructional practices, IEA studies have contributed greatly to current understanding of the educational process. PIRLS 2001 is a large international study of the reading literacy of young children around the world. It is designed to measure children's reading literacy achievement, to provide a baseline for future studies of trends in achievement, and to gather information about children's home and school experiences in learning to read.

In 1970, IEA conducted its first reading study, a study of reading comprehension in 15 countries.¹ This was followed in 1991 by the IEA Reading Literacy Study,² which studied student reading literacy achievement in 32 countries. PIRLS is the successor to these studies, and was influenced particularly

1 Thorndike, R.L. (1973). Reading comprehension in fifteen countries: An empirical study. *International studies in evaluation: Vol. 3*. Stockholm: Almqvist & Wiksell.

2 Elley, W.B. (Ed.). (1994). *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems*. Oxford, England: Elsevier Science Ltd.

by the Reading Literacy Study, which served as a starting point for its development. PIRLS 2001 is intended to be the first in a continuing five-year cycle of trend studies in reading literacy, and has been designed from the outset to monitor progress in reading achievement into the future. The IEA scheduled the PIRLS data collection for 2001 to coincide with the 10th anniversary of its 1991 Reading Literacy Study. In order to provide countries that participated in the 1991 Reading Literacy Study an opportunity to measure changes from 1991 to 2001, PIRLS 2001 also included an option to re-administer the 1991 reading literacy test in 2001, at the same time as the main PIRLS assessment. The results of this study, known as the Trends in IEA's Reading Literacy Study, are presented in a separate report.³

Participants in PIRLS

Thirty-five countries took part in the 2001 PIRLS assessment. Of these countries, nine participated in the Trends in IEA's Reading Literacy Study to measure changes between 1991 and 2001 in student performance as measured by the 1991 reading literacy test (see Exhibit A.1).

Developing the PIRLS Tests

The assessment framework and specifications⁴ for PIRLS was developed in collaboration with the PIRLS Reading Development Group (RDG) and with the assistance of the National Research Coordinators (NRCs) from the 35 countries participating in the study. The framework underwent several iterations in response to reviews and comments from the PIRLS countries and the reading research community, and embodies the ideas and interests of many individuals and organizations around the world. The IEA 1991 Reading Literacy Study served as the foundation for PIRLS, providing a basis for the PIRLS definition of reading literacy, and for establishing its framework and developing its assessment instruments. Although the 1991 study provided the groundwork for PIRLS, the PIRLS framework and instruments are new; reflecting the IEA's commitment to be forward-thinking, incorporating in PIRLS the latest approaches to measuring reading literacy.

3 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.

4 See 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.

Exhibit A.1: Countries Participating in PIRLS 2001 and the Trends in IEA's Reading Literacy Study

ISC

4th Grade
PIRLS 2001

Countries	PIRLS 2001	Trends Study
Argentina	●	
Belize	●	
Bulgaria	●	
Canada (O,Q)	●	
Colombia	●	
Cyprus	●	
Czech Republic	●	
England	●	
France	●	
Germany	●	
Greece	●	●
Hong Kong, SAR	●	
Hungary	●	●
Iceland	●	●
Iran, Islamic Rep. of	●	
Israel	●	
Italy	●	●
Kuwait	●	
Latvia	●	
Lithuania	●	
Macedonia, Rep. of	●	
Moldova, Rep. of	●	
Morocco	●	
Netherlands	●	
New Zealand	●	●
Norway	●	
Romania	●	
Russian Federation	●	
Scotland	●	
Singapore	●	●
Slovak Republic	●	
Slovenia	●	●
Sweden	●	●
Turkey	●	
United States	●	●

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2001.

The PIRLS assessment framework focuses on three aspects of reading literacy:

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

The first two aspects, the processes of comprehension and purposes for reading, form the basis of the written test of reading comprehension. The last aspect, the behaviors and attitudes, is addressed by the student questionnaire.

Readers construct meaning in different ways. They focus on and retrieve specific ideas, make inferences, interpret and integrate information and ideas, and examine or evaluate textual features. These four processes of comprehension were used in the PIRLS assessment to develop the comprehension questions for the passages presented to students. Across the assessment, a combination of questions, each dealing with one of the processes, enabled students to demonstrate a range of abilities and skills in constructing meaning from written texts.

Reading literacy is directly related to the reasons why people read. These reasons include reading for personal interest or pleasure, reading to participate in society, and reading to learn. For young readers, emphasis is placed on reading for interest or pleasure and reading to learn. The PIRLS assessment focused on the two purposes that account for most of the reading done by young students both in and out of school:

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

Because both types of reading are important at this age, the PIRLS assessment contained an equal proportion of materials assessing each purpose. Although the assessment distinguished between purposes for reading, the processes and strategies readers used for both purposes are perhaps more similar than different. The processes of comprehension and purposes for reading were the foundation for the PIRLS written assessment of reading comprehension.

Exhibit A.2 portrays the interaction of the two: each process is assessed within each purpose for reading.

The selection of the assessment passages and the development of the items and scoring guides were the result of an intensive process of collaboration, piloting, and review – spanning more than two years.⁵ In selecting the passages for PIRLS, every effort was made to minimize cultural bias. Potential stimulus passages were collected from as many countries as possible, and the final selection was based, in part, on the national and cultural representation of the entire set of assessment passages. Everything possible was done to ensure that the PIRLS assessment represented the curricula of the participating countries, and that the items did not exhibit bias towards or against particular countries. Draft passages and items were subjected to full-scale field testing before the instruments for the main data collection were finalized.⁶ The final version of the assessment was endorsed by the NRCs of the participating countries.

Exhibit A.3 shows the distribution of items by reading purpose and process category. There were 98 items in the assessment, approximately half of which were multiple-choice and half constructed-response. The constructed-response items required students to generate and write their own answers. Some items required short answers while others demanded a more elaborate response. In scoring the test, correct answers to most questions (including all those in multiple-choice format) were worth one point. However, responses to questions seeking more elaborate responses were evaluated for partial credit, with a fully-correct answer being awarded two or three points. Thus, the total number of score points available for analyses somewhat exceeds the number of items in the assessment. The student answer booklet provided an indication to the student of how many score points would be awarded for each answer, and how much writing was expected. Almost two-thirds of the score points came from constructed-response items.

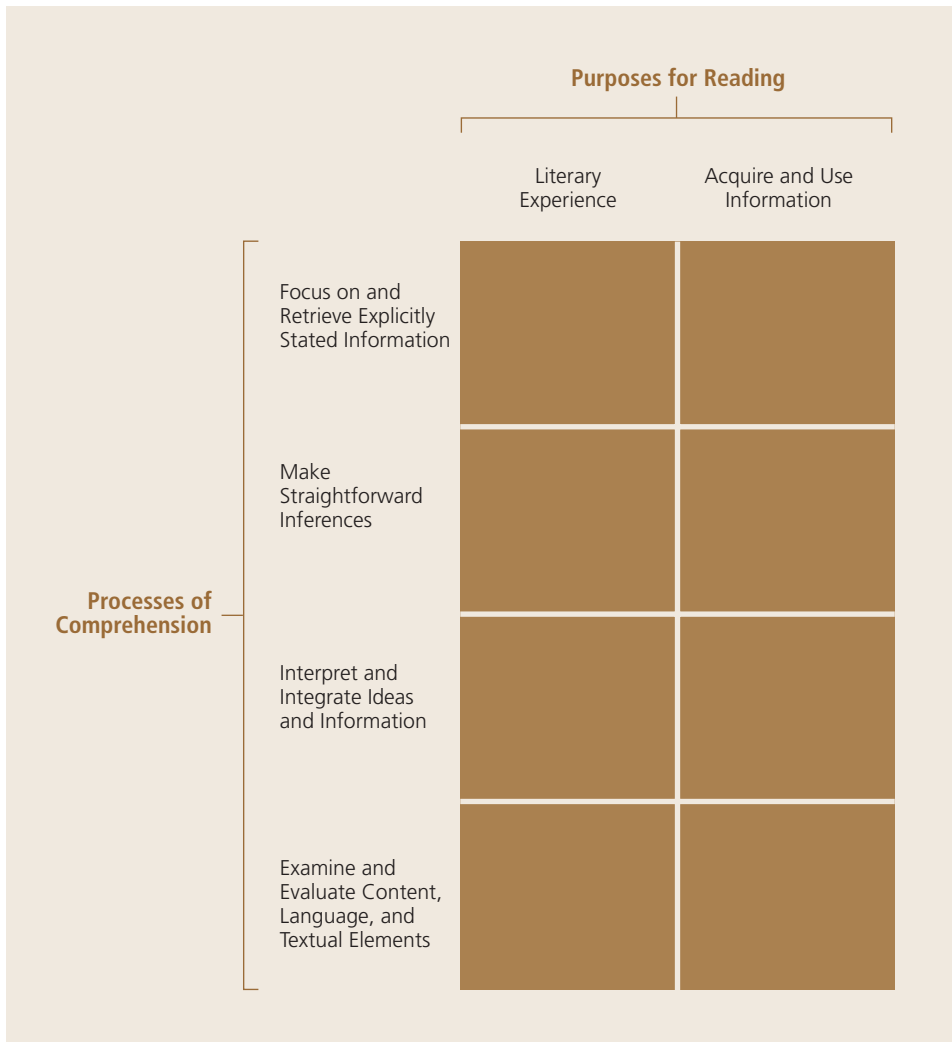
5 For a full discussion of the PIRLS 2001 test development effort, see Sainsbury, M. and Campbell, J.R. (2003). Developing the PIRLS reading assessment and scoring guides. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.

6 Approximately 48,000 students from almost 1,100 schools in 30 countries participated in the field test.

Exhibit A.2: Aspects of Reading Literacy

ISC 4th Grade
PIRLS 2001

PIRLS Reading Assessment – Reading Purposes and Processes



SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2001.

Exhibit A.3: Distribution of Items by Reading Purpose and Process Category**ISC** 4th Grade
PIRLS 2001

Reading Purpose	Percentage of Items	Total Number of Items	Number of Multiple-Choice Items	Number of Constructed-Response Items ¹	Number of Score Points ²
Literary Experience	52	51	25	26	66
Acquire and Use Information	48	47	21	26	67
Total	100	98	46	52	133

Reading Process Involved	Percentage of Items	Total Number of Items	Number of Multiple-Choice Items	Number of Constructed-Response Items ¹	Number of Score Points ²
Focus on and Retrieve Explicitly Stated Information and Ideas	26	25	15	10	29
Make Straightforward Inferences	28	27	14	13	31
Interpret and Integrate Ideas and Information	32	31	8	23	54
Examine and Evaluate Content, Language, and Textual Elements	15	15	9	6	19
Total	100	98	46	52	133

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2001.

1 Constructed-response items include both short-answer and extended-response types.

2 In scoring the tests, correct answers to most items were worth one point. However, responses to some constructed-response items were evaluated for partial credit with a fully correct answer awarded up to three points. Thus, the number of score points exceeds the number of items in the test.

PIRLS Test Design

Given the broad coverage goals of the PIRLS framework and its emphasis on the use of authentic texts, the specifications for the pool of reading passages and accompanying items included extensive testing time. The PIRLS RDG found that a valid assessment of two purposes for reading – reading for literary experience and reading to acquire and use information – required at least eight passages and items, four for each purpose. Students were given 40 minutes to complete a passage. With eight passages, a total assessment time would take up 320 minutes. While such an amount of assessment materials would provide good coverage of the reading material children meet in their everyday lives, it was an unreasonable expectation to administer the entire set of reading passages and test items to any one child. So as not to overburden the young children participating in PIRLS, the testing time was limited to 80 minutes (two passages) per student, with an additional 15–30 minutes allotted for a student questionnaire.

With eight reading passages in total, but just two to be given to any one student, passages and their accompanying items were assigned to student test booklets according to a matrix sampling plan. The eight passages were distributed across 10 booklets, two per booklet, so that passages were paired together in a booklet in as many different ways as possible. Each student booklet consisted of two 40-minute blocks of passages and items, which were accompanied by the Student Questionnaire. So as to present at least some passages in a more natural, authentic setting, two blocks (one literary and one informational) were presented in colorized, magazine-type format, with the questions appearing in a separate booklet. This booklet, Booklet 10, is referred to as the PIRLS “Reader.”

Background Questionnaires

The PIRLS questionnaires are grounded in a conceptual model relating reading outcomes – students’ reading literacy achievement and attitudes – to home, school, communal, and national contexts.⁷ PIRLS 2001 administered a broad array of questionnaires to collect data on these educational contexts.

7 See Kelly, D.L. (2003). Developing the PIRLS background questionnaires. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.

The students who were tested answered questions pertaining to their home and school experiences in learning to read – including instructional experiences, self-perception and attitudes towards reading, out-of-school reading habits, computer use, home literacy resources, and basic demographic information. Parents or caregivers of the sampled students responded to questions about the students' early reading experiences, child-parent literacy interactions, parents' reading habits and attitudes, home-school connections, and demographic and socioeconomic indicators. The teachers of the sampled students responded to questions about characteristics of the class tested, instructional activities for teaching reading, classroom resources, assessment practices, and about their education, training, and opportunities for professional development. The principals of schools responded to questions about enrollment and school characteristics, school organization for reading instruction, school staffing and resources, home-school connections, and the school environment.

Translation Verification

The PIRLS instruments were prepared in English and translated into 31 languages. Five countries administered the assessment in two languages, and seven countries administered one or more questionnaires in more than one language. The languages in which the test was administered most often were English (seven countries) and Arabic (three countries). In addition, it was sometimes necessary to modify the international versions for cultural reasons, even in the seven countries that tested in English. This process represented an enormous effort for the national centers, with many checks along the way. Before the translated instruments were used in schools, they were put through an exhaustive process of review and verification.⁸ The translation effort included: (1) developing explicit guidelines for translation and cultural adaptation; (2) translation of the instruments by the national centers in accordance with the guidelines – using two or more independent translations; (3) verification of translation quality by professional translators from an independent translation company; (4) corrections by national centers in accordance with the suggestions made; (5) verification by the PIRLS International Study Center

8 More details about the translation verification procedures can be found in Kelly, D.L., and Malak, B. (2003). Translating the PIRLS reading assessment and questionnaires. In M.O. Martin, I.V.S. Mullis, and A.M. Kennedy (Eds.), *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.

that the corrections were made; and (6) a series of statistical checks after the testing to detect items that did not perform comparably across countries.

For the participating countries, the bulk of the translation effort took place prior to the field test. After the field test, countries needed only to make any changes to the items or passages that resulted from analysis of the field test data. The PIRLS data-collection instruments were verified twice – the field test versions before the field test and the final versions before the main data collection. Countries, therefore, had the benefit of two careful reviews of their translations. They also had the benefit of diagnostic item statistics from the field test data analysis, which helped to identify mistranslations that could be corrected before the main data collection.

Sample Implementation and Participation Rates

PIRLS 2001 had as its target population students enrolled in the upper of the two adjacent grades that contained the largest proportion of 9-year-old students at the time of testing.⁹ Beyond the age criterion embedded in the above definition, the target grade should represent that point in the curriculum where students have essentially finished learning the basic reading skills and will focus more on “reading to learn” in the subsequent grades. Thus, the PIRLS 2001 target grade was expected to be fourth grade in most countries (some countries have students significantly older than nine years of age). Exhibit A.4 shows any differences in coverage between the international and national desired populations.

Selecting valid and efficient samples is critical to the quality and success of an international comparative study such as PIRLS. The accuracy of the survey results depends upon the quality of the sampling information available when planning the sample, and on the care with which the sampling activities are conducted. For PIRLS, NRCs worked on all phases of sampling in conjunction with staff from Statistics Canada. NRCs were trained in how to select the school and student samples, and in how to use the sampling software provided by the IEA Data Processing Center. In consultation with the PIRLS 2001 sampling referee (Keith Rust, Westat, Inc.), staff from Statistics Canada reviewed the national sampling plans, sampling data, sampling frames,

9 See Foy, P. and Joncas, M. (2003). PIRLS sampling design. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.

Exhibit A.4: Coverage of PIRLS Target Population

ISC
4th Grade
PIRLS 2001

Countries	International Desired Population		National Desired Population		
	Country Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Argentina	100%		3.7%	0.4%	4.1%
Belize	100%		0.8%	0.0%	0.8%
Bulgaria	100%		2.7%	0.0%	2.7%
Canada (O,Q)	60%	Provinces of Ontario and Quebec only	3.1%	2.2%	5.4%
Colombia	100%		3.2%	0.1%	3.3%
Cyprus	100%		0.0%	2.0%	2.0%
Czech Republic	100%		5.0%	0.0%	5.0%
England	100%		1.8%	3.9%	5.7%
France	100%		5.1%	0.3%	5.3%
Germany	100%		0.8%	1.0%	1.8%
Greece	100%		2.0%	5.3%	7.3%
Hong Kong, SAR	100%		2.8%	0.0%	2.8%
Hungary	100%		2.1%	0.0%	2.1%
Iceland	100%		1.8%	1.3%	3.1%
Iran, Islamic Rep. of	100%		0.5%	0.0%	0.5%
Israel	100%		16.5%	5.9%	22.4%
Italy	100%		0.0%	2.9%	2.9%
Kuwait	100%		0.0%	0.0%	0.0%
Latvia	100%		4.3%	0.3%	4.6%
Lithuania	90%	Lithuanian speaking students only	1.3%	2.5%	3.8%
Macedonia, Rep. of	100%		3.8%	0.4%	4.2%
Moldova, Rep. of	100%		0.5%	0.0%	0.5%
Morocco	100%		1.0%	0.0%	1.0%
Netherlands	100%		3.4%	0.3%	3.7%
New Zealand	100%		1.6%	1.7%	3.2%
Norway	100%		1.9%	0.8%	2.8%
Romania	100%		2.6%	1.9%	4.5%
Russian Federation	100%		2.8%	3.8%	6.6%
Scotland	100%		3.8%	0.8%	4.7%
Singapore	100%		1.3%	0.1%	1.4%
Slovak Republic	100%		1.4%	0.6%	2.0%
Slovenia	100%		0.0%	0.3%	0.3%
Sweden	100%		2.5%	2.5%	5.0%
Turkey	100%		3.9%	0.0%	3.9%
United States	100%		0.6%	4.7%	5.3%

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2001.

and sample selections. The sampling documentation was used by the International Study Center (in consultation with Statistics Canada and the sampling referee) to evaluate the quality of the samples.

The basic design of the sample used in PIRLS 2001 was a two-stage stratified cluster design. The first stage consisted of a sampling of schools, and the second stage of a sampling of intact classrooms from the target grade in the sampled schools. Most countries sampled 150 schools and one intact classroom from each school.¹⁰ Countries that selected large school samples included some countries with very large populations (such as the United States and the Russian Federation) as well as countries such as Canada, Germany, and Hungary which required accurate survey estimates for regions or provinces. Schools were selected with probability proportional to size, and classrooms with equal probabilities. Upon recommendation from Statistics Canada, some countries chose to sample more than one classroom per selected school.

Exhibits A.5 and A.6 present achieved sample sizes for schools and students, respectively. Exhibit A.7 shows the participation rates for schools, students, and overall – both with, and without, the use of replacement schools. For analysis and reporting, students' questionnaire data, along with questionnaire data from their parents, teachers, and school principals, were linked to students' achievement data. Exhibit A.8 shows the percentage of students with available student, parent, teacher, and principal questionnaire data.

Data Collection

Each participating country was responsible for carrying out all aspects of the data collection, using standardized procedures developed for the study. Training manuals were created for school coordinators and test administrators that explained procedures for receipt and distribution of materials, as well as for the activities related to the testing sessions. These manuals covered procedures for test security, standardized scripts to regulate directions and timing, rules for answering students' questions, and steps to ensure that identification on the test booklets and questionnaires corresponded to the information on the forms used to track students.

10 For further detail, see Joncas, M. (2003). PIRLS sampling weights and participation rates. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.

Exhibit A.5: School Sample Sizes

ISC
4th Grade
PIRLS 2001

Countries	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
Argentina	150	150	133	5	138
Belize	150	150	119	1	120
Bulgaria	177	176	170	0	170
Canada (O,Q)	387	387	359	13	372
Colombia	150	150	119	28	147
Cyprus	150	150	148	2	150
Czech Republic	150	148	135	6	141
England	150	150	88	43	131
France	150	150	140	5	145
Germany	216	215	209	2	211
Greece	170	170	133	12	145
Hong Kong, SAR	150	150	115	32	147
Hungary	220	220	216	0	216
Iceland	140	140	133	0	133
Iran, Islamic Rep. of	184	184	180	4	184
Israel	150	150	144	3	147
Italy	184	184	164	20	184
Kuwait	150	150	133	2	135
Latvia	148	147	133	8	141
Lithuania	150	150	84	62	146
Macedonia, Rep. of	150	150	145	1	146
Moldova, Rep. of	150	150	133	17	150
Morocco	158	158	117	0	117
Netherlands	150	150	80	54	134
New Zealand	156	156	144	12	156
Norway	162	160	119	17	136
Romania	150	150	144	0	144
Russian Federation	206	206	205	1	206
Scotland	150	150	113	5	118
Singapore	196	196	196	0	196
Slovak Republic	150	150	130	20	150
Slovenia	150	150	147	1	148
Sweden	150	149	142	4	146
Turkey	154	154	154	0	154
United States	200	200	125	49	174

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2001.

Exhibit A.6: Student Sample Sizes (Unweighted)

ISC
4th Grade
PIRLS 2001

Countries	Within-School Student Participation (Weighted Percentage)	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/School	Number of Students Excluded	Number of Eligible Students	Number of Students Absent	Number of Students Assessed
Argentina	91%	3769	132	13	3624	324	3300
Belize	94%	3137	32	0	3105	196	2909
Bulgaria	97%	3633	53	0	3580	120	3460
Canada (O,Q)	94%	9151	99	228	8824	571	8253
Colombia	96%	5582	225	5	5352	221	5131
Cyprus	97%	3149	2	63	3084	83	3001
Czech Republic	94%	3220	10	0	3210	188	3022
England	94%	3528	46	122	3360	204	3156
France	97%	3673	20	11	3642	104	3538
Germany	88%	8997	27	58	8912	1186	7726
Greece	97%	2718	0	151	2567	73	2494
Hong Kong, SAR	99%	5192	69	0	5123	73	5050
Hungary	97%	4819	14	0	4805	139	4666
Iceland	87%	4320	29	58	4233	557	3676
Iran, Islamic Rep. of	98%	7703	104	0	7599	169	7430
Israel	96%	4400	33	214	4153	180	3973
Italy	98%	3703	15	103	3585	83	3502
Kuwait	91%	7874	0	0	7874	741	7133
Latvia	93%	3266	8	11	3247	228	3019
Lithuania	85%	3114	7	72	3035	468	2567
Macedonia, Rep. of	97%	3904	42	14	3848	137	3711
Moldova, Rep. of	96%	3679	9	0	3670	137	3533
Morocco	93%	3452	35	0	3417	264	3153
Netherlands	98%	4256	11	14	4231	119	4112
New Zealand	96%	2720	68	53	2599	111	2488
Norway	92%	3784	25	26	3733	274	3459
Romania	97%	3744	23	2	3719	94	3625
Russian Federation	97%	4281	24	42	4215	122	4093
Scotland	95%	2912	20	26	2866	149	2717
Singapore	98%	7162	46	4	7112	110	7002
Slovak Republic	96%	4034	33	18	3983	176	3807
Slovenia	95%	3112	10	8	3094	142	2952
Sweden	93%	6678	38	145	6495	451	6044
Turkey	97%	5390	123	0	5267	142	5125
United States	96%	4091	55	121	3915	152	3763

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2001.

Exhibit A.7: Participation Rates (Weighted)

ISC
4th Grade
PIRLS 2001

Countries	School Participation		Student Participation	Overall Participation	
	Before Replacement	After Replacement		Before Replacement	After Replacement
Argentina	89%	92%	91%	81%	84%
Belize	80%	80%	94%	75%	75%
Bulgaria	97%	97%	97%	93%	93%
Canada (O,Q)	90%	97%	94%	85%	91%
Colombia	80%	98%	96%	76%	94%
Cyprus	98%	100%	97%	95%	97%
Czech Republic	90%	95%	94%	85%	90%
England	57%	87%	94%	54%	82%
France	93%	97%	97%	90%	94%
Germany	98%	98%	88%	86%	86%
Greece	78%	85%	97%	76%	82%
Hong Kong, SAR	73%	98%	99%	72%	97%
Hungary	98%	98%	97%	95%	95%
Iceland	95%	95%	87%	82%	82%
Iran, Islamic Rep. of	97%	100%	98%	95%	98%
Israel	96%	98%	96%	92%	94%
Italy	90%	100%	98%	88%	98%
Kuwait	87%	89%	91%	80%	81%
Latvia	89%	96%	93%	83%	89%
Lithuania	56%	97%	85%	47%	83%
Macedonia, Rep. of	97%	97%	97%	94%	94%
Moldova, Rep. of	84%	100%	96%	81%	96%
Morocco	74%	74%	93%	69%	69%
Netherlands	53%	89%	98%	52%	87%
New Zealand	94%	100%	96%	90%	96%
Norway	82%	89%	92%	76%	82%
Romania	96%	96%	97%	93%	93%
Russian Federation	100%	100%	97%	97%	97%
Scotland	76%	79%	95%	72%	74%
Singapore	100%	100%	98%	98%	98%
Slovak Republic	88%	100%	96%	84%	96%
Slovenia	98%	99%	95%	94%	94%
Sweden	97%	99%	93%	90%	92%
Turkey	100%	100%	97%	97%	97%
United States	61%	86%	96%	59%	83%

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2001.

Exhibit A.8: Percentage of Students with Any Available Student, Parent, Teacher, and Principal Questionnaire Data
ISC 4th Grade
PIRLS 2001

Countries	Percent of Student with Any Available Data			
	Student Questionnaire	Parent Questionnaire	Teacher Questionnaire	Principal Questionnaire
Argentina	96	71	96	98
Belize	95	67	87	89
Bulgaria	100	96	98	99
Canada (O,Q)	100	83	98	97
Colombia	100	90	97	99
Cyprus	99	68	98	89
Czech Republic	100	89	100	99
England	100	55	94	96
France	99	92	95	97
Germany	100	89	96	95
Greece	100	88	95	94
Hong Kong, SAR	98	95	97	97
Hungary	100	96	97	97
Iceland	99	84	93	85
Iran, Islamic Rep. of	100	98	100	100
Israel	100	47	97	97
Italy	100	97	100	100
Kuwait	80	82	90	95
Latvia	100	96	100	100
Lithuania	100	98	99	99
Macedonia, Rep. of	100	77	97	98
Moldova, Rep. of	99	99	99	100
Morocco	98	0	91	95
Netherlands	99	66	88	88
New Zealand	99	84	97	96
Norway	99	91	99	100
Romania	100	98	98	98
Russian Federation	100	99	100	100
Scotland	99	63	95	92
Singapore	100	99	100	100
Slovak Republic	99	97	100	100
Slovenia	100	98	99	98
Sweden	98	92	95	98
Turkey	99	99	100	100
United States	100	0	99	96

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2001.

Each country was responsible for conducting quality control procedures and describing this effort in the Survey Activities report documenting procedures used in the study. In addition, the International Study Center considered it essential to independently monitor compliance with standardized procedures. NRCs were asked to nominate a person, unconnected with their national center, to serve as quality control monitors (QCMs) for their countries. The International Study Center developed manuals for the quality control monitors and, in a two-day training session, briefed them about PIRLS, the responsibilities of the national centers in conducting the study, and their own roles and responsibilities. Monitors from 33 countries attended the training session conducted by the International Study Center staff. In countries where the data collection schedule made it impossible for one quality control monitor to visit all the sampled schools, monitors who attended the training session were asked to recruit other monitors as necessary, in order to allow for efficiency in the coverage of the territory and testing timetable. In all, 71 quality control monitors participated.¹¹ They interviewed NRCs about data collection plans and procedures, and visited a sample of 15 schools in each country, where they observed testing sessions and interviewed school coordinators.¹² All together, quality control monitors visited observed testing sessions and interviewed school coordinators in 475 schools from 33 countries.

The results of the interviews indicate that, in general, NRCs had prepared well for data collection and – despite the heavy demands of the schedule and shortages of resources – were able to conduct the data collection efficiently and professionally. Similarly, the PIRLS test appeared to have been administered in compliance with international procedures – including the activities before the testing session, along with school-level activities related to receiving, distributing, and returning material from national centers.

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. To ensure reliable scoring procedures based on the PIRLS rubrics, the International Study Center prepared detailed

11 Operational constraints did not permit QCM visits to be conducted in Argentina or Iceland.

12 Steps taken to ensure high-quality data collection in PIRLS are described in detail in Gonzalez, E.J., & Kennedy, A.M. (2003). Quality control in the PIRLS data collection. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.

guides containing the 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 in scoring the constructed-response items. The training sessions were designed to help representatives of national centers, who would then be responsible for training personnel in their own countries to apply the scoring rubrics reliably.

To gather and document empirical information about the within-country agreement among scorers, PIRLS arranged to have systematic sub-samples of at least 200 students' responses to each item scored independently by two readers. Exhibit A.9 shows the average range of the within-country exact percent of agreement between scorers on the free-response items. Scoring reliability within countries was high – the percentage of exact agreement, on average, across countries, was 93 percent.

To monitor the consistency with which the scoring rubrics were applied across countries, PIRLS collected from the countries that administered PIRLS in English a sample of 200 student responses to 25 constructed-response questions from four of the assessment passages. This set of 5,000 student responses was then sent to each country having scorers proficient in English, and were all scored independently by two of these scorers. Each of these responses was scored by 55 scorers from the countries that participated.¹³ Making all possible comparisons among scorers gave 1,485 comparisons for each student response to each item, and 297,000 total comparisons when aggregated across all 200 student responses to that item. Agreement across countries was defined in terms of the percentage of these comparisons that were in exact agreement. Exhibit A.10 shows this percentage of exact agreement for each of the 25 items. As shown in this exhibit, the percentage of agreement averaged across the 25 items was 85 percent.

Test Reliability

Exhibit A.11 displays the reading test reliability coefficient for each country. This coefficient is the median KR-20 reliability across the nine test booklets

¹³ Scorers proficient in English were available in 28 of the 35 PIRLS countries. Only one English-proficient scorer was available in Macedonia. In the Russian Federation, resources permitted only half of the English-language responses to be scored.

and the PIRLS Reader. Median reliabilities ranged from 0.83 in The Netherlands to 0.91 in England, Israel, Macedonia, New Zealand, Romania, and Singapore. The international median (0.88) is the median of the reliability coefficients for all countries.

Data Processing

To ensure the availability of comparable, high-quality data for analysis, PIRLS took rigorous quality control steps to create the international database.¹⁴ 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 in Hamburg for creation of the international database. Upon arrival at the Data Processing Center, the data underwent an exhaustive cleaning process. This involved 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.¹⁵ In general, the items exhibited very good psychometric properties in all countries. On only two occasions was an item deleted for a country; once because of a translation error in the student booklet, and once because of a misinterpretation of a scoring rubric.

IRT Scaling and Data Analysis

The general approach to reporting the PIRLS achievement data was based primarily on item response theory (IRT) scaling methods.¹⁶ Student reading achievement was summarized using a family of 2- and 3-parameter IRT models

14 These steps are detailed in Itzlinger, U., & Schwippert, K. (2003). Creating and checking the PIRLS database. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.

15 See Mullis, I.V.S., Martin, M.O., & Kennedy, A.M. (2003). Reviewing the PIRLS item statistics. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.

16 For a detailed description of the PIRLS scaling, see Gonzalez, E.J. (2003). Scaling the PIRLS reading assessment data. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.

Exhibit A.9: PIRLS Within-Country Constructed-Response Scoring Reliability Data
ISC 4th Grade
PIRLS 2001

Countries	Correctness Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent of Agreement	
		Minimum	Maximum
Argentina	86	71	95
Belize	92	86	97
Bulgaria	83	60	99
Canada (O,Q)	87	66	99
Colombia	83	65	100
Cyprus	96	86	100
Czech Republic	97	82	100
England	96	81	100
France	96	87	100
Germany	89	71	100
Greece	98	92	100
Hong Kong, SAR	88	61	97
Hungary	94	80	100
Iceland	86	70	99
Iran, Islamic Rep. of	95	90	99
Israel	91	83	97
Italy	94	68	100
Kuwait	–	–	–
Latvia	92	64	99
Lithuania	88	68	100
Macedonia, Rep. of	94	85	98
Moldova, Rep. of	94	83	99
Morocco	–	–	–
Netherlands	90	67	100
New Zealand	97	89	100
Norway	92	81	99
Romania	94	76	100
Russian Federation	98	91	100
Scotland	93	76	100
Singapore	99	98	100
Slovak Republic	99	99	100
Slovenia	92	67	100
Sweden	94	86	100
Turkey	99	98	100
United States	97	89	100
International Avg.	93	79	99

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2001.

A dash (–) indicates data are not available.

Exhibit A.10: PIRLS Cross-Country Constructed-Response Scoring Reliability

ISC 4th Grade
PIRLS 2001

Purpose	Item Label ¹	Total Valid Comparisons ²	Exact Percent Agreement
Literary Experience	Unreleased C01	275496	99%
	Unreleased C02	275444	89%
	Unreleased C03	275548	93%
	Unreleased C06	275341	98%
	Unreleased C08	275496	92%
	Unreleased C10	275548	66%
	Unreleased C11	275444	72%
	Hare H03	275600	90%
	Hare H04	275393	93%
	Hare H07	275444	79%
	Hare H08	275086	84%
	Hare H09	275236	84%
Hare H10	273661	73%	

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2001.

Acquire and Use Information	Unreleased A01	296892	96%
	Unreleased A03	296676	98%
	Unreleased A04	296676	90%
	Unreleased A07	296892	87%
	Unreleased A08	296623	80%
	Unreleased A09	296784	81%
	Unreleased A11	296191	80%
	Pufflings N07	274724	78%
	Pufflings N08	274724	83%
	Pufflings N10	273947	84%
	Pufflings N12	274673	76%
	Pufflings N13	274621	73%

Average Percent Agreement 85%

1 See Appendix C for item descriptions and scoring guides.

2 Values for items differ slightly due to a small number of missing responses.

Exhibit A.11 : Cronbach's Alpha Reliability Coefficient – PIRLS 2001**ISC** 4th Grade
PIRLS 2001

Countries	Reliability Coefficient ¹
Argentina	0.90
Belize	0.87
Bulgaria	0.89
Canada (O,Q)	0.87
Colombia	0.87
Cyprus	0.90
Czech Republic	0.85
England	0.91
France	0.87
Germany	0.87
Greece	0.88
Hong Kong, SAR	0.85
Hungary	0.87
Iceland	0.89
Iran, Islamic Rep.	0.89
Israel	0.91
Italy	0.87
Kuwait	0.86
Latvia	0.85
Lithuania	0.85
Macedonia, Rep. of	0.91
Moldova, Rep. of	0.87
Morocco	0.90
Netherlands	0.83
New Zealand	0.91
Norway	0.89
Romania	0.91
Russian Federation	0.86
Scotland	0.90
Singapore	0.91
Slovak Republic	0.88
Slovenia	0.88
Sweden	0.85
Turkey	0.89
United States	0.90
International Median	0.88

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2001.

1 The reliability coefficient for each country is the median Cronbach's alpha reliability across the ten test booklets.

for dichotomously-scored items (right or wrong), and generalized partial credit models for items with two or three available score points. The IRT scaling method produces a score by averaging the responses of each student to the items that he or she took which takes into account the difficulty and discriminating power of each item. The methodology used in PIRLS included refinements enabling reliable scores to be produced even though individual students responded to just two of the eight assessment passages. 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. Exhibit A.12 presents the Pearson correlation coefficient indicating the linear relationship between the two reading purposes in each of the PIRLS countries.

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 (Booklets 1 through 9, or the PIRLS Reader). 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 was set at 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.

To allow more accurate estimation of summary statistics for student subpopulations, the PIRLS scaling made use of plausible-value technology, whereby five separate estimates of each student's score were generated on each scale – based on the student's responses to the items in the student's booklet, and on the student's background characteristics. The five score estimates are known as "plausible values," and the variability between them encapsulates the uncertainty inherent in the score estimation process.

Exhibit A.12: Correlation Between Reading for Literary Purposes and Reading for Informational Purposes
ISC 4th Grade
PIRLS 2001

Countries	Pearson Correlation Coefficient
Argentina	0.81
Belize	0.85
Bulgaria	0.85
Canada (O,Q)	0.82
Colombia	0.83
Cyprus	0.87
Czech Republic	0.81
England	0.88
France	0.81
Germany	0.87
Greece	0.82
Hong Kong, SAR	0.84
Hungary	0.84
Iceland	0.81
Iran, Islamic Rep. of	0.83
Israel	0.88
Italy	0.81
Kuwait	0.80
Latvia	0.85
Lithuania	0.81
Macedonia, Rep. of	0.90
Moldova, Rep. of	0.78
Morocco	0.81
Netherlands	0.79
New Zealand	0.88
Norway	0.87
Romania	0.88
Russian Federation	0.76
Scotland	0.88
Singapore	0.94
Slovak Republic	0.85
Slovenia	0.88
Sweden	0.83
Turkey	0.83
United States	0.88
International Med.	0.84

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2001.

Estimating Sampling Error

Because the statistics presented in this report are estimates of national performance based on samples of students – rather than on the values that could be calculated if every student in every country had answered every question – it is important to have measures for the degree of uncertainty of the estimates. The jackknife procedure was used to estimate the standard error associated with each statistic presented in this report.¹⁷ The jackknife standard errors also include an error component due to variation between the five plausible values generated for each student. The use of confidence intervals (based on the standard errors) provides a way to make inferences about the population means and proportions in a manner that reflects the uncertainty associated with the sample estimates. An estimated sample statistic plus or minus two standard errors represents a 95 percent confidence interval for the corresponding population result.

Setting International Benchmarks of Student Achievement

To facilitate reporting of student reading achievement at a variety of performance levels, PIRLS identified four international benchmarks of student achievement. These benchmarks are the points on the PIRLS reading scale that separate the 10 percent of students located on top of the distribution, the top 25 percent of students, the top 50 percent, and the bottom 25 percent. The percentage of students in each country meeting or exceeding the international benchmarks is reported. The benchmarks correspond to the 90th, 75th, 50th, and 25th percentiles of the international distribution of achievement. When computing these percentiles, sampling weights were applied so that each country contributed as many students to the distribution as there were students in the target population in the country. That is, each country's contribution to setting the international benchmarks was proportional to the estimated population enrolled at the fourth grade.

In order to interpret the PIRLS scale scores and analyze achievement at the international benchmarks, PIRLS conducted a scale anchoring analysis to describe achievement of students at those four points on the scale.

17 Procedures for computing jackknifed standard errors are presented in Gonzalez, E.J., & Kennedy, A.M. (2003). Statistical analysis and reporting of the PIRLS data. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College.

Scale anchoring is a way of describing students' performance at different points on a scale, in terms of the kind of reading they can do and the level of comprehension they exhibit. It involves a statistical component, in which items that discriminate between successive points on the scale are identified, and a judgmental component in which subject matter experts examine the items and generalize to students' knowledge and understandings.¹⁸ In PIRLS, the Reading Development Group (RDG) worked with the Reading Coordinator and PIRLS staff to describe student reading at the international benchmarks.

¹⁸ The scale-anchoring procedure is described fully in Gonzalez, E.J., & Kennedy, A.M. (2003). Statistical analysis and reporting of the PIRLS data. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2001 technical report*. Chestnut Hill, MA: Boston College. An application of the procedure to the 1995 TIMSS data may be found in Kelly, D.L., Mullis, I.V.S., & Martin, M.O. (2000). *Profiles of student achievement in mathematics at the TIMSS international benchmarks: U.S. performance and standards in an international context*. Chestnut Hill, MA: Boston College.

