Appendix A

Supplementary Information About PIRLS 2006 Procedures

Participants in PIRLS 2006

Building on two earlier IEA studies of reading achievement, the 1970 study of reading comprehension in 15 countries[1] and the 1991 Reading Literacy Study[2] in 32 countries, the PIRLS assessment of student achievement in reading literacy at the fourth grade is an integral component of IEA's ongoing program of studies in the core subjects of mathematics, science, and reading.[3] Beginning with PIRLS 2001,[4] PIRLS was designed from the outset to monitor progress in reading achievement on a regular 5-year cycle. PIRLS 2006, the second study in this continuing cycle, was designed to measure children's reading literacy achievement, to provide information on changes in achievement since 2001, and to add to the store of knowledge about children's home and school experiences in learning to read.

Forty countries, including Belgium with 2 education systems and Canada with 5 provinces, participated in the 2006 PIRLS assessment for a total of 45 participants. Of these, 26 countries and 2 provinces had trend data from PIRLS 2001.[5] Participating in PIRLS for the first time in 2006 were 13 countries (counting Belgium as one country) and 3 provinces (see Exhibit A.1).

The PIRLS 2006 Test Instruments

Across the PIRLS 2006 assessment, the questions on the reading passages enabled students to demonstrate a range of abilities and skills in constructing meaning from written texts. An important innovation in PIRLS 2006 was the ability to report achievement results according to reading comprehension processes, in addition to reading purposes. In PIRLS 2001, achievement results were reported for overall reading comprehension and separately by literary and informational purposes, but not by process of comprehension. Subsequently, research has indicated that it would be possible also to report comprehension processes separately by combining the retrieval and straightforward inferencing processes to make one scale and the interpreting and integrating and examining and evaluating processes to make another.[6]

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.

3 Mathematics and science are assessed at fourth and eighth grades by IEA's TIMSS (Trends in International Mathematics and Science Study).

4 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.

5 Although Kuwait participated in PIRLS 2001, the data were not considered comparable for measuring trends, and so Kuwait does not appear in any trend exhibits.

6 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). *International achievement in the processes of reading comprehension: results from PIRLS 2001 in 35 countries*. (2004). Chestnut Hill, MA: Boston College.

Exhibit A.1 Countries Participating in PIRLS 2006 and 2001

PIRLS 2006
4th Grade

Countries	2006	2001
Argentina		●
Austria	●	
Belgium (Flemish)	●	
Belgium (French)	●	
Belize		●
Bulgaria	●	●
Canada, Alberta	●	
Canada, British Columbia	●	
Canada, Nova Scotia	●	
Canada, Ontario	●	●
Canada, Quebec	●	●
Chinese Taipei	●	
Colombia		●
Cyprus		●
Czech Republic		●
Denmark	●	
England	●	●
France	●	●
Georgia	●	
Germany	●	●
Greece		●
Hong Kong SAR	●	●
Hungary	●	●
Iceland	●	●
Indonesia	●	
Iran, Islamic Rep. of	●	●
Israel	●	●
Italy	●	●
[1] Kuwait	●	
Latvia	●	●
Lithuania	●	●
Luxembourg	●	
Macedonia, Rep. of	●	●
Moldova, Rep. of	●	●
Morocco	●	●
Netherlands	●	●
New Zealand	●	●
Norway	●	●
Poland	●	
Qatar	●	
Romania	●	●
Russian Federation	●	●
Scotland	●	●
Singapore	●	●
Slovak Republic	●	●
Slovenia	●	●
South Africa	●	
Spain	●	
Sweden	●	●
Trinidad and Tobago	●	
Turkey		●
United States	●	●

Indicates country participation in that testing cycle ●

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

1 Although Kuwait participated in PIRLS 2001, the data were not considered comparable for measuring trends.

However, this necessitated increasing the number of passages and items in the assessment from 8 to 10 to ensure that there were sufficient items for the process scales.

Half of the passages and items from the 2001 assessment, two literary and two informational, have been published with the international report so that readers could appreciate the nature of the PIRLS reading tasks, and half were kept secure to serve as a basis for linking to the PIRLS 2006 assessment. The four secure passages and items (two literary and two informational) were available for use again in 2006. However, in addition, it was necessary to develop 6 new passages and items to replace the released passages and to expand the scope of the assessment from 8 to 10 passages.

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.[7] Draft passages and items were subjected to full-scale field testing before the instruments for the main data collection were finalized.[8] The final version of the assessment was endorsed by the NRCs of the participating countries.

Exhibit A.2 shows the distribution of the PIRLS 2006 test items by reading purpose and process category. There were 126 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 (167) 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. About 60 percent of the score points came from constructed-response items. Of the 126 items, 49 were trend items, that is, items from

7 For a full discussion of the PIRLS 2006 test development effort, see Kennedy, A.M. & Sainsbury, M. (2007). Developing the PIRLS 2006 reading assessment and scoring guides. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2006 technical report*. Chestnut Hill, MA: Boston College.

8 Approximately 50,000 students from almost 1,200 schools in 42 countries participated in the field test.

Exhibit A.2 **Distribution of Items by Reading Purpose and Process Category** **PIRLS 2006** 4th Grade

Items in the PIRLS 2006 Assessment				
Reading Purpose	**Total Number of Items**	**Number of Multiple-choice Items**	**Number of Constructed-response Items**	**Total Number Score Points**
Literary Experience	64	34	30	85
Acquire and Use Information	62	30	32	82
Total	126	64	62	167

Reading Process	**Percentage of Items**	**Total Number of Items**	**Number of Multiple-choice Items**	**Number of Constructed-response Items**	**Number of Score Points**
Focus on and Retrieve Explicitly Stated Information and Ideas	22	31	19	12	36
Make Straightforward Inferences	28	43	29	14	47
Interpret and Integrate Ideas and Information	37	34	6	28	61
Examine and Evaluate Content, Language, and Textual Elements	14	18	10	8	23
Total	100	126	64	62	167

Trend Items in the PIRLS 2006 Assessment (Items also used in PIRLS 2001)				
Reading Purpose	**Total Number of Items**	**Number of Multiple-choice Items**	**Number of Constructed-response Items**	**Total Number Score Points**
Literary Experience	26	13	13	33
Acquire and Use Information	23	10	13	33
Total	49	23	26	66

Reading Process	**Total Number of Items**	**Number of Multiple-choice Items**	**Number of Constructed-response Items**	**Number of Score Points**
Focus on and Retrieve Explicitly Stated Information and Ideas	12	5	7	15
Make Straightforward Inferences	18	10	8	21
Interpret and Integrate Ideas and Information	12	3	10	22
Examine and Evaluate Content, language, and Textual Elements	7	5	2	8
Total	49	23	27	66

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

the 4 passages that were retained from 2001 to serve as the basis of the link between the 2001 and 2006 assessments.

PIRLS Testing Time and Booklet Design

Given the broad coverage goals of the PIRLS 2006 framework and its emphasis on the use of authentic texts, the passages and accompanying items required extensive testing time. Students were given 40 minutes to complete a passage. With 10 passages, it would have taken 400 minutes to administer the entire assessment to a single child. However, so as not to overburden the relatively young children participating in PIRLS, and in line with the practice in 2001, the testing time was limited to 80 minutes (two passages) per student, with an additional 15–30 minutes allotted for a student questionnaire.

With 10 reading passages in total, but just 2 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 10 passages were distributed across 13 booklets, 2 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. 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 13, is referred to as the PIRLS "Reader."[9]

Translation Verification

The PIRLS 2006 instruments were prepared in English and translated into 45 languages (Exhibit A.3). Although most countries administered the assessment in just one language, nine countries and the five Canadian provinces administered it in two languages, and Spain administered the assessment in its five official languages: Spanish (Castilian), Catalonian, Galician, Basque, and Valencian. Most impressive of all, South Africa administered the PIRLS 2006 assessment in 11 languages: English, Afrikaans,

9 The PIRLS 2006 test booklet design is described in 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.

isiZulu, isiXhosa, Sepedi, Sesotho, Setswana, isiNdebele, Siswati, Tshivenda, and Xitsonga.

The test was administered most often in English (seven countries, counting the Canadian provinces once), with Arabic and French second (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.

The translation process represented an enormous effort for the national centers, with many checks along the way, including an exhaustive process of review and verification.[10]

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 the IEA Secretariat using professional translators from an independent translation company, (4) corrections by national centers in accordance with the suggestions made, (5) verification by the TIMSS & PIRLS International Study Center that the corrections were made and that the layout of the instruments corresponded to the international standard, 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 changes to the items or passages that resulted from analysis of the field test data.

The translations of the PIRLS 2006 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.

10 More details about the translation verification procedures can be found in Malak, B. & Trong, K. L. (2007). Translating the PIRLS 2006 reading assessment and questionnaires. In M.O. Martin, I.V.S. Mullis, and A.M. Kennedy (Eds.), *PIRLS 2006 technical report*. Chestnut Hill, MA: Boston College.

Exhibit A.3 Languages of Instruction and Testing

PIRLS 2006
4th Grade

Country	Language(s) of Instruction	Language(s) of Testing
Austria	German	German
Belgium (Flemish)	Dutch	Dutch
Belgium (French)	French	French
Bulgaria	Bulgarian	Bulgarian
Canada, Alberta	English, French	English, French
Canada, British Columbia	English, French	English, French
Canada, Nova Scotia	English, French	English, French
Canada, Ontario	English, French	English, French
Canada, Quebec	English, French, Aboriginal languages	English, French
Chinese Taipei	Mandarin	Chinese Mandarin
Denmark	Danish	Danish
England	English	English
France	French	French
Georgia	Georgian	Georgian
Germany	German	German
Hong Kong SAR	Chinese	Modern Standard Chinese
Hungary	Hungarian	Hungarian
Iceland	Icelandic	Icelandic
Indonesia	Indonesian	Indonesian
Iran, Islamic Rep. of	Farsi	Farsi
Israel	Hebrew, Arabic	Hebrew, Arabic
Italy	Italian, French, German, Ladin, and Slovenian	Italian
Kuwait	Arabic and local dialects	Arabic
Latvia	Latvian, Russian	Latvian, Russian
Lithuania	Lithuanian, Russian, Polish	Lithuanian
Luxembourg	Luxembourgish, French, German	German
Macedonia, Rep. of	Macedonian, Albanian, Turkish and Serbian	Macedonian, Albanian
Moldova, Rep. of	Romanian, Russian	Romanian, Russian
Morocco	–	Arabic
Netherlands	Dutch	Dutch
New Zealand	English, Māori	English, Maori
Norway	Bokmål, Nynorsk, Sámi	Bokmål, Nynorsk
Qatar	Arabic, English	Arabic
Poland	Polish	Polish
Romania	Romanian, Hungarian	Romanian, Hungarian
Russian Federation	Russian	Russian
Scotland	English, Gaelic	English
Singapore	English, Malay, Chinese (Mandarin), Tamil	English
Slovak Republic	Slovak, Hungarian	Slovak, Hungarian
Slovenia	Slovenian	Slovenian
South Africa	Afrikaans, English, isiZulu, isiXhosa, Sepedi, Sesotho, Setswana, isiNdebele, SiSwati, Tshivenda, Xitsonga	Afrikaans, English, isiZulu, isiXhosa, Sepedi, Sesotho, Setswana, isiNdebele, SiSwati, Tshivenda, Xitsonga
Spain	Spanish (Castilian), Catalonian, Galician, Basque, Valencian	Spanish (Castilian), Catalonian, Galician, Basque, Valencian
Sweden	Swedish	Swedish
Trinidad and Tobago	English	English
United States	English	English

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

A dash (–) indicates comparable data are not available.

This table lists the languages most frequently used for instruction. More detailed information for each country is available in the *PIRLS 2006 Encyclopedia*.

Sample Implementation and Participation Rates

PIRLS 2006 had as its target population students enrolled in the fourth grade of formal schooling, counting from the first year of primary school as defined by UNESCO's International Standard Classification for Education (ISCED).[11] According to the ISCED classification, Level 1 corresponds to primary education or the first stage of basic education, and the first year of Level 1 should mark the beginning of formal instruction in reading, writing, and mathematics. Accordingly, the fourth year of Level 1 should be fourth grade in most countries. To avoid testing very young children, however, PIRLS has a policy that the average age of children in the grade tested should not be below 9.5 years old.

The PIRLS 2006 assessment was administered to carefully drawn random samples of students from the target population in each country. Because the accuracy of the PIRLS results depends on the quality of the national samples, the PIRLS team went to great lengths to work with participating countries to ensure efficient sampling design and implementation.

For PIRLS 2006, national research coordinators worked on all phases of sampling in conjunction with staff from Statistics Canada. National coordinators were trained in how to select the school and student samples, and in how to use the *WinW3S* within-school sampling software provided by the IEA Data Processing and Research Center. In consultation with the PIRLS 2006 sampling referee (Keith Rust, Westat, Inc.), staff from Statistics Canada reviewed the national sampling plans, sampling data, sampling frames, and sample selections. The sampling documentation was used by the TIMSS & PIRLS International Study Center (in consultation with Statistics Canada and the sampling referee) to evaluate the quality of the samples.

In a few situations where it was not possible to test the entire internationally desired population (i.e., all students enrolled in the fourth grade), countries were permitted to define a national desired population that excluded part of the internationally desired population. Exhibit A.4 shows any differences in coverage between the international and national desired populations. Almost all participants achieved 100% coverage,

11 UNESCO Institute for Statistics. (1999). *Operational manual for ISCED-1997.*

the exceptions being Georgia (tested only Georgian-speaking students), Lithuania (only Lithuanian-speaking students), and Moldova (did not include the Predniestrian republic).

Within the desired population, countries could define a population that excluded a small percentage (less than 5%) of certain kinds of schools or students that would be very difficult or resource intensive to test (e.g., schools for students with special needs or schools that were very small or located in remote rural areas). Almost all countries kept their excluded students below the 5 percent limit, except for Bulgaria, Denmark, Georgia, the Russian Federation, and the United States, which just exceeded this figure, and Israel, which excluded more that 20 percent of its fourth-grade student population.

The basic design of the sample used in PIRLS 2006 was a two-stage stratified cluster design.[12] The first stage was a sampling of schools, and the second stage a sampling of intact classrooms from the target grade in the sampled schools. Schools were selected with probability proportional to size, and classrooms with equal probabilities. Most countries sampled 150 schools and one or two intact classrooms from each school.[13] This approach was designed to yield a representative sample of at least 4,000 students in each country.

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. Most countries achieved the minimum acceptable participation rates—85 percent of both the schools and students, or a combined rate (the product of school and student participation) of 75 percent—although Belgium (Flemish), the Netherlands, Scotland, and the United States did so only after including replacement schools and have been annotated in the exhibits of this report. Norway had overall participation rates after including replacement schools of just below 75 percent (71%) and has been annotated accordingly.

Because an important goal of the PIRLS 2006 assessment was to measure changes in fourth-grade students' reading achievement since 2001, it is important to track any changes in population composition and coverage since then that might be related to student achievement. Exhibit A.8 presents,

12 See Joncas, M. (2007). PIRLS 2006 sampling design. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), PIRLS *2006 technical report*. Chestnut Hill, MA: Boston College.

13 For further detail, see Joncas, M. (2007). PIRLS 2006 sampling weights and participation rates. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2006 technical report*. Chestnut Hill, MA: Boston College.

for each country, four attributes of the populations sampled in 2001 and 2006: number of years of formal schooling, average student age, the score on the United Nations Development Programme's (UNDP) human development index, and the percentage of students in the national desired population excluded from the assessment. Most countries and provinces were very similar with regard to these attributes across the two assessments, although it is noteworthy than the Russian Federation and Slovenia underwent structural changes in the age at which children enter schools that are reflected in their samples. In 2001, the Russian sample contained third-grade students from some regions and fourth-grade students from others, whereas all students were in fourth grade in 2006. Slovenia is in transition toward having all children begin school at an earlier age so that they all will have four years of primary schooling instead of three years, as was the case in 2001. However, the transition was not complete in 2006.

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.9 shows the percentage of students with available student, parent, teacher, and principal questionnaire data. Although the vast majority of students in PIRLS 2006 were taught by a single teacher, there were some students in some countries taught by more than one teacher. The percentage of students in each country taught by one, two, or three teachers is presented in Exhibit A.10. Only Scotland, Sweden, and the Canadian province of British Columbia had more than 10 percent of students with more than one teacher—21 percent, 13 percent, and 13 percent, respectively. If a student had more than one teacher, the student's data record was replicated so that there were as many student records as there were teacher records. Then each teacher record was merged with one of the student records. So as not to over count the student in analyses, the sampling weight for each student record was divided by the number of records that the student had. For example, if a student had two teachers, the student's record was replicated so that there were two records, and the first was merged with one teacher record and the second with the other teacher record. Each of the two student records was given half of the original student's weight.

Exhibit A.4 Coverage of PIRLS Target Population

PIRLS 2006
4th Grade

Countries	International Desired Population		National Desired Population		
	Country Coverage	Notes on Coverage	School-level Exclusions	Within-sample Exclusions	Overall Exclusions
Austria	100%		1.4%	3.8%	5.1%
Belgium (Flemish)	100%		6.1%	1.1%	7.1%
Belgium (French)	100%		3.7%	0.3%	3.9%
Bulgaria	100%		2.2%	4.3%	6.4%
Canada, Alberta	100%		2.0%	5.2%	7.1%
Canada, British Columbia	100%		2.2%	5.5%	7.6%
Canada, Nova Scotia	100%		0.2%	3.8%	4.0%
Canada, Ontario	100%		1.6%	6.8%	8.3%
Canada, Quebec	100%		2.4%	1.2%	3.6%
Chinese Taipei	100%		1.8%	1.1%	2.9%
Denmark	100%		0.5%	5.7%	6.2%
England	100%		1.6%	0.9%	2.4%
France	100%		3.4%	0.4%	3.8%
Georgia	80%	Students taught in Georgian	2.4%	5.0%	7.3%
Germany	100%		0.4%	0.3%	0.7%
Hong Kong SAR	100%		3.0%	0.9%	3.9%
Hungary	100%		2.3%	1.4%	3.7%
Iceland	100%		1.3%	2.5%	3.8%
Indonesia	100%		3.2%	0.0%	3.2%
Iran, Islamic Rep. of	100%		2.9%	0.9%	3.8%
Israel	100%		17.5%	6.1%	22.5%
Italy	100%		0.1%	5.2%	5.3%
Kuwait	100%		0.3%	0.0%	0.3%
Latvia	100%		4.3%	0.5%	4.7%
Lithuania	93%	Students taught in Lithuanian	0.9%	4.2%	5.1%
Luxembourg	100%		0.9%	3.0%	3.9%
Macedonia, Rep. of	100%		4.6%	0.3%	4.9%
Moldova, Rep. of	91%	Moldova less Predniestrian – Moldovan Republic	0.6%	0.0%	0.6%
Morocco	100%		1.1%	0.0%	1.1%
Netherlands	100%		3.5%	0.1%	3.6%
New Zealand	100%		1.4%	3.9%	5.3%
Norway	100%		1.0%	2.8%	3.8%
Poland	100%		0.9%	4.2%	5.1%
Qatar	100%		0.7%	0.7%	1.4%
Romania	100%		2.4%	0.0%	2.4%
Russian Federation	100%		6.8%	1.0%	7.7%
Scotland	100%		1.4%	0.9%	2.3%
Singapore	100%		0.9%	0.0%	0.9%
Slovak Republic	100%		1.8%	1.9%	3.6%
Slovenia	100%		0.2%	0.5%	0.8%
South Africa	100%		4.2%	0.1%	4.3%
Spain	100%		1.3%	4.0%	5.3%
Sweden	100%		2.4%	1.5%	3.9%
Trinidad and Tobago	100%		0.7%	0.0%	0.7%
United States	100%		3.2%	2.8%	5.9%

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

Exhibit A.5 **School Sample Sizes**

PIRLS 2006
4th Grade

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
Austria	160	158	158	0	158
Belgium (Flemish)	150	149	102	35	137
Belgium (French)	150	150	129	21	150
Bulgaria	150	147	130	13	143
Canada, Alberta	150	150	150	0	150
Canada, British Columbia	150	150	147	1	148
Canada, Nova Scotia	201	201	200	1	201
Canada, Ontario	200	198	173	7	180
Canada, Quebec	200	194	185	0	185
Chinese Taipei	150	150	147	3	150
Denmark	150	146	128	17	145
England	150	150	129	19	148
France	175	175	164	5	169
Georgia	152	149	139	10	149
Germany	410	407	397	8	405
Hong Kong SAR	150	144	130	14	144
Hungary	150	149	147	2	149
Iceland	136	131	128	0	128
Indonesia	170	168	166	2	168
Iran, Islamic Rep. of	240	236	235	1	236
Israel	150	149	146	3	149
Italy	150	150	136	14	150
Kuwait	150	150	149	0	149
Latvia	150	150	145	2	147
Lithuania	150	146	144	2	146
Luxembourg	183	178	178	0	178
Macedonia, Rep. of	150	150	149	1	150
Moldova, Rep. of	150	150	148	2	150
Morocco	160	160	156	3	159
Netherlands	150	150	104	35	139
New Zealand	250	250	220	23	243
Norway	178	177	118	17	135
Poland	150	148	147	1	148
Qatar	123	119	119	0	119
Romania	150	147	146	0	146
Russian Federation	232	232	232	0	232
Scotland	150	150	101	29	130
Singapore	178	178	178	0	178
Slovak Republic	174	171	155	12	167
Slovenia	150	150	140	5	145
South Africa	441	410	391	6	397
Spain	152	152	149	3	152
Sweden	150	147	147	0	147
Trinidad and Tobago	150	149	147	0	147
United States	222	214	120	63	183

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

Exhibit A.6 **Student Sample Sizes**

PIRLS 2006
4th Grade

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 Students Eligible	Number of Students Absent	Number of Students Assessed
Austria	98%	5431	24	208	5199	132	5067
Belgium (Flemish)	99%	4608	10	47	4551	72	4479
Belgium (French)	95%	4810	19	14	4777	225	4552
Bulgaria	97%	4156	37	135	3984	121	3863
Canada, Alberta	96%	4773	79	250	4444	201	4243
Canada, British Columbia	95%	4663	68	244	4351	201	4150
Canada, Nova Scotia	96%	4884	79	189	4616	180	4436
Canada, Ontario	97%	4436	40	252	4144	156	3988
Canada, Quebec	84%	4639	50	99	4490	742	3748
Chinese Taipei	99%	4746	62	55	4629	40	4589
Denmark	97%	4349	51	154	4144	143	4001
England	93%	4492	117	38	4337	301	4036
France	98%	4558	55	16	4487	83	4404
Georgia	98%	4837	120	209	4508	106	4402
Germany	94%	8395	49	44	8302	403	7899
Hong Kong SAR	97%	4917	25	34	4858	146	4712
Hungary	97%	4265	17	46	4202	134	4068
Iceland	91%	4200	47	102	4051	378	3673
Indonesia	98%	4981	99	0	4882	108	4774
Iran, Islamic Rep. of	99%	5609	122	22	5465	54	5411
Israel	93%	4378	5	179	4194	286	3908
Italy	97%	3882	31	153	3698	117	3581
Kuwait	89%	4467	0	0	4467	509	3958
Latvia	94%	4469	14	17	4438	276	4162
Lithuania	92%	5400	67	183	5150	449	4701
Luxembourg	99%	5342	15	158	5169	68	5101
Macedonia, Rep. of	96%	4209	33	11	4165	163	4002
Moldova, Rep. of	95%	4281	32	0	4249	213	4036
Morocco	95%	3444	43	0	3401	152	3249
Netherlands	97%	4366	63	5	4298	142	4156
New Zealand	96%	6872	130	196	6546	290	6256
Norway	87%	4570	27	134	4409	572	3837
Poland	95%	5410	21	232	5157	303	4854
Qatar	94%	7490	305	47	7138	458	6680
Romania	98%	4463	97	0	4366	93	4273
Russian Federation	97%	4911	20	35	4856	136	4720
Scotland	94%	4123	66	41	4016	241	3775
Singapore	95%	6760	67	0	6693	303	6390
Slovak Republic	96%	5741	34	105	5602	222	5380
Slovenia	96%	5596	12	27	5557	220	5337
South Africa	92%	16144	305	28	15811	1154	14657
Spain	97%	4391	12	143	4236	142	4094
Sweden	96%	4653	33	33	4587	193	4394
Trinidad and Tobago	95%	4237	77	0	4160	209	3951
United States	96%	5761	160	159	5442	252	5190

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

Exhibit A.7 Participation Rates (Weighted)

PIRLS 2006
4th Grade

Countries	School Participation		Classroom Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Austria	100%	100%	99%	98%	97%	97%
Belgium (Flemish)	69%	92%	100%	99%	68%	91%
Belgium (French)	85%	100%	100%	95%	81%	95%
Bulgaria	88%	97%	100%	97%	85%	94%
Canada, Alberta	100%	100%	100%	96%	96%	96%
Canada, British Columbia	98%	99%	100%	95%	93%	94%
Canada, Nova Scotia	99%	100%	100%	96%	96%	96%
Canada, Ontario	88%	90%	100%	97%	85%	87%
Canada, Quebec	96%	96%	100%	84%	81%	81%
Chinese Taipei	98%	100%	100%	99%	97%	99%
Denmark	89%	99%	100%	97%	86%	96%
England	86%	99%	100%	93%	80%	92%
France	94%	97%	100%	98%	92%	95%
Georgia	94%	100%	100%	98%	93%	98%
Germany	97%	99%	100%	94%	90%	92%
Hong Kong SAR	91%	100%	100%	97%	89%	97%
Hungary	99%	100%	100%	97%	96%	97%
Iceland	99%	99%	100%	91%	90%	90%
Indonesia	99%	100%	100%	98%	97%	98%
Iran, Islamic Rep. of	100%	100%	100%	99%	99%	99%
Israel	98%	100%	100%	93%	91%	93%
Italy	91%	100%	100%	97%	88%	97%
Kuwait	99%	99%	99%	89%	88%	88%
Latvia	97%	98%	100%	94%	91%	92%
Lithuania	99%	100%	100%	92%	90%	92%
Luxembourg	100%	100%	100%	99%	99%	99%
Macedonia, Rep. of	100%	100%	100%	96%	96%	96%
Moldova, Rep. of	98%	100%	100%	95%	93%	95%
Morocco	98%	99%	100%	95%	93%	94%
Netherlands	70%	93%	100%	97%	67%	90%
New Zealand	92%	99%	100%	96%	88%	95%
Norway	68%	82%	100%	87%	58%	71%
Poland	99%	100%	100%	95%	94%	95%
Qatar	100%	100%	100%	94%	94%	94%
Romania	99%	99%	100%	98%	97%	97%
Russian Federation	100%	100%	100%	97%	97%	97%
Scotland	69%	87%	100%	94%	65%	81%
Singapore	100%	100%	100%	95%	95%	95%
Slovak Republic	93%	98%	100%	96%	89%	94%
Slovenia	93%	97%	100%	96%	90%	93%
South Africa	94%	96%	100%	92%	86%	88%
Spain	99%	100%	100%	97%	95%	97%
Sweden	100%	100%	100%	96%	96%	96%
Trinidad and Tobago	99%	99%	100%	95%	94%	94%
United States	57%	86%	100%	96%	54%	82%

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

Exhibit A.8 **Trends in Student Populations**

PIRLS 2006
4th Grade

Country	Years of Formal Schooling		Average Age		Human Development Index		Overall Exclusion Rate	
	2006	2001	2006	2001	2006[1]	2001[2]	2006	2001
Bulgaria	4	4	10.9	10.9	0.816	0.772	6.4%	2.7%
Canada, Ontario	4	4	9.8	9.9	0.950	0.936	8.3%	6.6%
Canada, Quebec	4	4	10.1	10.2	0.950	0.936	3.6%	3.3%
England	5	5	10.3	10.2	0.940	0.923	2.4%	5.7%
France	4	4	10.0	10.1	0.942	0.924	3.8%	5.3%
Germany	4	4	10.5	10.5	0.932	0.921	0.7%	1.8%
Hong Kong SAR	4	4	10.0	10.2	0.927	0.880	3.9%	2.8%
Hungary	4	4	10.7	10.7	0.869	0.829	3.7%	2.1%
Iceland	4	4	9.8	9.7	0.960	0.932	3.8%	3.1%
Iran	4	4	10.2	10.4	0.746	0.714	3.8%	0.5%
Israel	4	4	10.1	10.0	0.927	0.893	22.5%	22.4%
Italy	4	4	9.7	9.8	0.940	0.909	5.3%	2.9%
Kuwait	4	4	9.8	9.9	0.871	0.818	0.3%	0.0%
Latvia	4	4	11.0	11.0	0.845	0.791	4.7%	4.6%
Lithuania	4	4	10.7	10.9	0.857	0.803	5.1%	3.8%
Macedonia	4	4	10.6	10.7	0.796	0.766	4.9%	4.2%
Moldova	4	4	10.9	10.8	0.694	0.699	0.6%	0.5%
Morocco	4	4	10.8	11.2	0.640	0.596	1.1%	1.0%
Netherlands	4	4	10.3	10.3	0.947	0.931	3.6%	3.7%
New Zealand	5	5	10.0	10.1	0.936	0.913	5.3%	3.2%
Norway	4	4	9.8	10.0	0.965	0.939	3.8%	2.8%
Romania	4	4	10.9	11.1	0.805	0.772	2.4%	4.5%
Russian Federation	4	3 or 4	10.8	10.3	0.797	0.775	7.7%	6.6%
Scotland	5	5	9.9	9.8	0.940	0.923	2.3%	4.7%
Singapore	4	4	10.4	10.1	0.916	0.876	0.9%	1.4%
Slovak Republic	4	4	10.4	10.3	0.856	0.831	3.6%	2.0%
Slovenia	3 or 4	3	9.9	9.8	0.910	0.874	0.8%	0.3%
Sweden	4	4	10.9	10.8	0.951	0.936	3.9%	5.0%
United States	4	4	10.1	10.2	0.948	0.934	5.9%	5.3%

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

1 Taken from the United Nations Development Programme's *Human Development Report 2006*, p. 283–286

2 Taken from the United Nations Development Programme's *Human Development Report 2001*, p. 141–144

Exhibit A.9 **Percentage of Students with Any Available Student, Parent, Teacher, and Principal Questionnaire Data**

PIRLS 2006
4th Grade

Countries	Percent of Student with Any Available Data			
	Student Questionnaire	Parent Questionnaire	Teacher Questionnaire	Principal Questionnaire
Austria	100	96	100	100
Belgium (Flemish)	100	97	99	96
Belgium (French)	100	90	96	87
Bulgaria	99	96	97	96
Canada, Alberta	99	80	99	99
Canada, British Columbia	99	77	80	88
Canada, Nova Scotia	99	91	91	96
Canada, Ontario	100	90	99	95
Canada, Quebec	99	90	95	92
Chinese Taipei	100	97	99	100
Denmark	100	94	95	95
England	100	46	91	83
France	99	92	98	96
Georgia	100	98	99	100
Germany	96	87	94	96
Hong Kong SAR	98	98	99	97
Hungary	100	90	99	98
Iceland	99	76	90	91
Indonesia	100	99	100	100
Iran, Islamic Rep. of	100	99	100	100
Israel	99	63	98	98
Italy	100	97	100	100
Kuwait	97	75	90	95
Latvia	100	95	98	99
Lithuania	100	98	100	99
[1] Luxembourg	100	93	99	0
Macedonia, Rep. of	98	97	95	84
Moldova, Rep. of	100	97	96	97
Morocco	100	98	99	74
Netherlands	100	67	90	85
New Zealand	99	65	96	96
Norway	98	93	98	95
Poland	100	98	100	100
Qatar	99	72	81	92
Romania	100	98	100	99
Russian Federation	100	99	100	100
Scotland	100	52	88	80
Singapore	100	98	100	100
Slovak Republic	100	97	99	100
Slovenia	100	95	100	99
South Africa	99	90	95	99
Spain	100	62	96	91
Sweden	100	94	95	99
Trinidad and Tobago	99	89	97	98
[2] United States	100	0	99	99

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

1 Primary schools in Luxembourg do not have principals.

2 All countries except the United States administered the parents' questionnaire. In exhibits presenting data from this questionnaire, the United States has dashes (–).

Exhibit A.10 **Percentage of Students with One or More Teachers**

PIRLS 2006
4th Grade

Country	Percent with One Teacher	Percent with Two Teachers	Percent with Three Teachers
Austria	99	1	0
Belgium (Flemish)	100	0	0
Belgium (French)	100	0	0
Bulgaria	100	0	0
Canada, Alberta	95	5	0
Canada, British Columbia	87	13	0
Canada, Nova Scotia	92	8	0
Canada, Ontario	100	0	0
Canada, Quebec	100	0	0
Chinese Taipei	100	0	0
Denmark	96	4	0
England	91	6	3
France	100	0	0
Georgia	99	1	0
Germany	97	3	0
Hong Kong SAR	100	0	0
Hungary	100	0	0
Iceland	92	8	0
Indonesia	100	0	0
Iran, Islamic Rep. of	100	0	0
Israel	100	0	0
Italy	100	0	0
Kuwait	100	0	0
Latvia	100	0	0
Lithuania	100	0	0
Luxembourg	90	8	2
Macedonia, Rep. of	100	0	0
Moldova, Rep. of	100	0	0
Morocco	100	0	0
Netherlands	99	1	0
New Zealand	95	5	0
Norway	100	0	0
Poland	100	0	0
Qatar	100	0	0
Romania	100	0	0
Russian Federation	100	0	0
Scotland	79	21	1
Singapore	100	0	0
Slovak Republic	100	0	0
Slovenia	100	0	0
South Africa	100	0	0
Spain	100	0	0
Sweden	86	13	1
Trinidad and Tobago	100	0	0
United States	100	0	0

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

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.

Each country was responsible for conducting quality control procedures and describing this effort in the online Survey Activities Report. In addition, the TIMSS & PIRLS International Study Center considered it essential to independently monitor compliance with standardized procedures.[14] To implement the independent monitoring program, the IEA Secretariat asked National Research Coordinators to nominate persons, unconnected with their national centers, to serve as quality control monitors (QCMS) for their countries. All countries and provinces participated in the program of quality control school visits.

The TIMSS & PIRLS International Study Center developed manuals for the quality control monitors and, in a 2-day training session, staff briefed the monitors about PIRLS 2006, the responsibilities of the national centers in conducting the study, and their roles and responsibilities as quality control monitors. The training session, jointly conducted by the TIMSS & PIRLS International Study Center and the IEA Secretariat, was attended by 42 quality control monitors. 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, 103 quality control monitors and assistants participated in the program, visiting a sample of 15 schools in each country, where they

14 Steps taken to ensure high-quality data collection in PIRLS 2006 are described in detail in Johansone, I. & Kennedy, A.M. (2007). Quality assurance in the PIRLS 2006 data collection. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2006 technical report*. Chestnut Hill, MA: Boston College.

observed testing sessions and interviewed school coordinators. Altogether, quality control monitors observed testing sessions and interviewed school coordinators in 669 schools from across all 45 PIRLS 2006 participants.

National Research Coordinators' comments in the Survey Activities Report indicate that, in general, national centers 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, based on quality control monitors observations of the testing sessions, there is evidence that the PIRLS 2006 test was 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 2006 implemented procedures for reliably evaluating student responses within and across countries. The TIMSS & PIRLS International Study Center prepared detailed scoring guides containing the classification categories and explanations of how to implement the classifications, together with example student responses for the various categories. These scoring guides, along with training packets containing extensive examples of student responses for practice in applying the guides, were used as a basis for intensive training in scoring the constructed-response items. The TIMSS & PIRLS International Study Center conducted scoring training sessions for the PIRLS 2006 participants in conjunction with both the field test and the PIRLS 2006 assessment. The training sessions were designed for representatives from national centers, who would then be responsible for training personnel in their own countries to apply the scoring guides reliably.

To gather and document empirical information about the within-country agreement among scorers, PIRLS arranged to have systematic subsamples of at least 200 students' responses to each item scored independently by two

readers. Exhibit A.11 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.

PIRLS 2006 also took steps to ensure that those constructed-response items from the 2001 passages that were used in 2006 were scored in the same way in both assessments. In anticipation of this, countries that participated in PIRLS 2001 sent samples of scored student booklets from their 2001 assessment to the IEA Data Processing and Research Center (DPC), where they were digitally scanned and incorporated into custom-built presentation software for use in 2006. On average, the software contained about 5,000 student responses for each country. After being trained in using the scoring guides for these items, scorers scored half of the student responses, using the scoring software supplied by the IEA DPC. The software then produced reports on their scoring accuracy for these student responses. Scorers with less than 85 percent exact agreement with the scores assigned to the responses in 2001 were retrained before proceeding. Exhibit A.12 provides the average percentage agreement across items for the scores given in 2001 and in 2006 for each participant. Agreement between 2001 and 2006 was generally high—90 percent exact agreement on average across countries.[15]

To monitor the consistency with which the scoring rubrics were applied across countries, PIRLS 2006 collected from the countries that administered PIRLS in English a sample of 200 student responses to 23 constructed-response questions from four of the assessment passages. This set of 4,600 student responses was then sent to each country having scorers proficient in English, to be scored independently by two of these scorers. Each of these responses was scored by 62 scorers from across the countries that participated.[16] Making all possible comparisons among scorers gave 1,891 comparisons for each student response to each item, and 378,200 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.13 shows this percentage of exact agreement for each of the 23 items. As shown in this exhibit, the percentage of agreement averaged across the 23 items was 87 percent.

15 A number of participants were unable to complete the trend-scoring reliability task, because of software difficulties or because it was not possible to scan their 2001 student booklets.

16 Scorers proficient in English were available in 33 of the PIRLS 2006 countries. In some countries, more than two scorers shared the scoring effort.

Exhibit A.11 PIRLS Within-country Constructed-response Scoring Reliability Data

PIRLS 2006
4th Grade

Countries	Correctness Score Agreement		
	Average of Exact Percent Agreement Across Items	Range of Exact Percent of Agreement	
		Minimum	Maximum
Austria	95	80	100
Belgium (Flemish)	90	73	99
Belgium (French)	97	90	100
Bulgaria	98	94	100
Canada, Alberta	91	67	100
Canada, British Columbia	92	70	100
Canada, Nova Scotia	93	84	100
Canada, Ontario	94	80	100
Canada, Quebec	95	87	100
Chinese Taipei	95	78	100
Denmark	97	90	100
England	98	93	100
France	89	69	100
Georgia	85	65	98
Germany	89	76	99
Hong Kong SAR	96	85	100
Hungary	98	89	100
Iceland	95	88	99
Indonesia	95	76	100
Iran, Islamic Rep. of	93	83	99
Israel	91	80	98
Italy	95	85	100
Kuwait	86	80	95
Latvia	90	78	100
Lithuania	97	91	100
Luxembourg	94	82	100
Macedonia, Rep. of	88	78	96
Moldova, Rep. of	99	97	100
Morocco	89	71	97
Netherlands	99	93	100
New Zealand	93	80	98
Norway	83	66	97
Poland	97	93	100
Qatar	97	93	99
Romania	99	96	100
Russian Federation	99	97	100
Scotland	97	89	100
Singapore	98	94	100
Slovak Republic	96	88	100
Slovenia	98	92	100
South Africa	82	63	92
Spain	81	61	96
Sweden	92	72	100
Trinidad and Tobago	93	71	100
United States	93	82	100
International Avg.	93	82	99

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

Exhibit A.12 **PIRLS 2006 Trend Scoring Reliability (2001–2006) for the Constructed–response Items** **PIRLS 2006** 4th Grade

Countries	Average Exact Percent Agreement Across Items
Bulgaria	–
Canada, Ontario	–
Canada, Quebec	–
England	89
France	90
Germany	88
Hong Kong SAR	93
Hungary	91
Iceland	–
Iran, Islamic Rep. of	92
Israel	96
Italy	91
Latvia	84
Lithuania	92
Macedonia, Rep. of	81
Moldova, Rep. of	–
Morocco	–
Netherlands	93
New Zealand	90
Norway	90
Romania	–
Russian Federation	–
Scotland	88
Singapore	88
Slovak Republic	92
Slovenia	–
Sweden	89
United States	93
International Avg.	90

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

A dash (–) indicates data are not available.

Exhibit A.13 PIRLS Cross-country Constructed-response Scoring Reliability

PIRLS 2006
4th Grade

Purpose	Item Label[1]	Total Valid Comparisons[2]	Exact Percent Agreement
Literary Experience	Flowers F06C	377504	91%
	Flowers F07C	377957	80%
	Flowers F08C	375960	92%
	Flowers F09C	378078	93%
	Flowers F10C	376869	97%
	Flowers F12C	375684	63%
	Unbelievable Night U05C	377224	99%
	Unbelievable Night U06C	377385	93%
	Unbelievable Night U08C	378078	76%
	Unbelievable Night U10C	377453	96%
	Unbelievable Night U12C	377302	87%
Acquire and Use Information	Antartica A01C	378200	95%
	Antartica A03C	378139	98%
	Antartica A04C	377542	89%
	Antartica A07C	378139	88%
	Antartica A08C	377722	80%
	Antartica A09C	377370	83%
	Antartica A11C	377363	81%
	Day Hiking N02C	377897	91%
	Day Hiking N03C	378139	94%
	Day Hiking N08C	376927	92%
	Day Hiking N11C	377773	77%
	Day Hiking N12C	330146	76%
		Average Percent Agreement	87%

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

1 See Appendix D for item descriptions and scoring guides.

2 The number of comparisons varies across items because not all scorers scored all items.

Test Reliability

Exhibit A.14 displays the reading test reliability coefficient for each country. This coefficient is the median KR-20 reliability across the 12 test booklets and the PIRLS Reader. Reliabilities were generally high. Almost all countries had reliabilities between 0.8 and 0.9, and eight countries—Bulgaria, England, Israel, Macedonia, New Zealand, Romania, South Africa, and Trinidad and Tobago—had reliabilities of 0.9 or greater. The median of the reliability coefficients across all countries was 0.88.

Data Processing

To ensure the availability of comparable, high-quality data for analysis, PIRLS 2006 took rigorous quality control steps to create the international database.[17] 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 and Research Center (DPC) in Hamburg for creation of the international database. Upon arrival at the IEA DPC, the data underwent an exhaustive quality-control process. This involved an iterative procedure of checking, editing, and rechecking 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 DPC, the TIMSS & PIRLS 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 DPC, the TIMSS & PIRLS International Study Center reviewed item statistics for each achievement item in each country to identify poorly performing items.[18] In general, the items exhibited very good psychometric properties in all countries. However, as a result of the item review, one item was deleted from the achievement scaling for all countries because of a

17 These steps are detailed in Barth, J., & Neuschmidt, O. (2007). Creating and checking the PIRLS 2006 database. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2006 technical report*. Chestnut Hill, MA: Boston College.

18 See Martin, M.O., Kennedy, A.M. , & Trong, K. L. (2004). Reviewing the PIRLS 2006 item statistics. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2006 technical report*. Chestnut Hill, MA: Boston College.

Exhibit A.14 **Cronbach's Alpha Reliability Coefficient – Overall Reading**

PIRLS 2006
4th Grade

Countries	Reliability Coefficient
Austria	0.86
Belgium (Flemish)	0.83
Belgium (French)	0.86
Bulgaria	0.90
Canada, Alberta	0.86
Canada, British Columbia	0.86
Canada, Nova Scotia	0.88
Canada, Ontario	0.87
Canada, Quebec	0.86
Chinese Taipei	0.86
Denmark	0.87
England	0.91
France	0.86
Georgia	0.87
Germany	0.86
Hong Kong SAR	0.82
Hungary	0.88
Iceland	0.88
Indonesia	0.81
Iran, Islamic Rep. of	0.88
Israel	0.91
Italy	0.87
Kuwait	0.85
Latvia	0.86
Lithuania	0.83
Luxembourg	0.88
Macedonia, Rep. of	0.91
Moldova, Rep. of	0.86
Morocco	0.87
Netherlands	0.81
New Zealand	0.91
Norway	0.86
Poland	0.89
Qatar	0.84
Romania	0.90
Russian Federation	0.88
Scotland	0.89
Singapore	0.89
Slovak Republic	0.89
Slovenia	0.88
South Africa	0.92
Spain	0.88
Sweden	0.85
Trinidad and Tobago	0.91
United States	0.88
International Median	0.88

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

The reliability coefficient for each country is the median KR-20 reliability across the 13 test booklets.

problem with the scoring guide, and on a few occasions, items were deleted for individual countries because of translation errors.

IRT Scaling and Data Analysis

The primary approach to reporting the PIRLS 2006 achievement data was based on item response theory (IRT) scaling methods.[19] Student reading achievement was summarized using a family of 2- and 3-parameter IRT models for dichotomously-scored items (right or wrong), and generalized partial credit models for items with 2 or 3 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 in a way that takes into account the difficulty and discriminating power of each item.

A notable feature of IRT scaling is that it is capable of estimating a student's score on an assessment even if the student has not responded to all of the items in the assessment pool. This characteristic of IRT scaling makes it particularly appropriate for PIRLS, where each individual student completed a single test booklet, comprising just two of the 10 passages in the PIRLS 2006 assessment (approximately 17 score points per passage). The PARSCALE[20] program was used to estimate the IRT model parameters.

Although IRT methods are well suited to the PIRLS design, like other measurement approaches they provide the most reliable results when based on large numbers of items. Because individual students respond to relatively few items, particularly on the reading purposes and processes of comprehension scales, PIRLS uses a process known as "conditioning" to improve the reliability of the achievement measurement. The conditioning process combines students' responses to the items they were administered with information about the students' background characteristics to construct a distribution of achievement for each student that is conditional on the student's responses to the administered items and on the student's background characteristics.

To provide student scores that may be used in analyses, PIRLS uses the achievement distribution to predict or impute the achievement of each student conditional on his or her item responses and background

19 For a detailed description of the PIRLS 2006 scaling, see Foy, P., Galia, J., & Li, Isaac. (2007). Scaling the PIRLS 2006 reading assessment data. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2006 technical report*. Chestnut Hill, MA: Boston College.

20 Muraki, E., & Bock, R.D. (1997). PARSCALE: IRT item analysis and test scoring for rating-scale data [Computer software and manual]. Chicago: Scientific Software.

characteristics. These imputed scores, or "plausible values," are used as scale scores in analyses to create the exhibits in this report. To quantify any error in the imputation process, PIRLS generates five plausible values for each student and conducts all analyses five times. The average of the results of the five analyses is taken as the best estimate of the statistic in question, and the difference between them reflects the imputation error. PIRLS uses the MGROUP[21] program developed by Educational Testing Service to implement the conditioning and generate plausible values.

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.

The PIRLS reading achievement scales were designed to reliably measure student achievement on the same scale for both the 2001 and 2006 assessments. The metric of the scales was established originally with the 2001 assessment. Treating all countries participating in PIRLS 2001 equally, the PIRLS scale average across those countries was set at 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 preserve the metric of the original 2001 scale, the 2006 assessment was first scaled using all of the items from both 2001 and 2006 and all students from countries that participated in both 2001 and 2006. Although the items from the four passages used in both assessments were the foundation for linking the two sets of assessment data, all items from 2001 and 2006 were included in this scaling. Having established the characteristics of the scale, scores were computed for students from countries that participated in 2006 but not in 2001.

Achievement scales were produced for each of the two reading purposes (reading for literary experience and reading for information) and for two processes of comprehension (retrieving and straightforward inferencing, and interpreting, integrating, and evaluating), as well as for reading overall. Exhibit A.15 presents the Pearson correlation coefficient indicating the linear

21 Sheehan, K.M. (1985). M-GROUP: Estimation of group effects in multivariate models [Computer software and manual]. Princeton, NJ: Educational Testing Service.

Exhibit A.15 Correlation Between Two Scales for Purposes and Two Scales for Processes for Reading

PIRLS 2006 4th Grade

Countries	Pearson Correlation Coefficient	
	Between Reading for Literary Experience and Reading for Information	Between Retrieval and Straightforward Inferencing and Interpreting, Integrating, and Evaluating
Austria	0.90	0.92
Belgium (Flemish)	0.82	0.91
Belgium (French)	0.88	0.92
Bulgaria	0.83	0.92
Canada, Alberta	0.81	0.92
Canada, British Columbia	0.83	0.90
Canada, Nova Scotia	0.86	0.93
Canada, Ontario	0.83	0.91
Canada, Quebec	0.81	0.88
Chinese Taipei	0.85	0.91
Denmark	0.87	0.92
England	0.89	0.94
France	0.86	0.91
Georgia	0.81	0.91
Germany	0.90	0.93
Hong Kong SAR	0.77	0.87
Hungary	0.87	0.92
Iceland	0.86	0.92
Indonesia	0.76	0.88
Iran, Islamic Rep. of	0.88	0.93
Israel	0.89	0.96
Italy	0.82	0.89
Kuwait	0.79	+
Latvia	0.84	0.89
Lithuania	0.85	0.90
Luxembourg	0.90	0.92
Macedonia, Rep. of	0.90	0.95
Moldova, Rep. of	0.80	0.88
Morocco	0.82	+
Netherlands	0.79	0.89
New Zealand	0.91	0.96
Norway	0.82	0.90
Poland	0.89	0.92
Qatar	0.88	+
Romania	0.88	0.95
Russian Federation	0.88	0.93
Scotland	0.89	0.93
Singapore	0.90	0.96
Slovak Republic	0.91	0.94
Slovenia	0.88	0.94
South Africa	0.93	+
Spain	0.82	0.92
Sweden	0.89	0.91
Trinidad and Tobago	0.92	0.95
United States	0.87	0.96
International Median	0.88	0.92

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

A plus (+) sign indicates average achievement could not be accurately estimated on the interpreting, integrating, and evaluating scale.

relationship between the two reading purposes in each of the PIRLS 2006 countries and between the two processes of comprehension. Across countries, the median correlation between reading for literary experience and reading for information was 0.88, and the median correlation between the comprehension processes retrieving and simple inferencing and interpreting, integrating, and evaluating was 0.92.

To facilitate comparisons of countries' relative performance on the two reading purposes (i.e., do students perform relatively better reading for literary experience or reading for information?) and on the two comprehension processes (again, do students perform relatively better on one process than the other?) PIRLS 2006 set all scales to have the same scale average and standard deviation—500 with a standard deviation of 100. This means that any existing differences in the overall difficulty of the items comprising each of the scales are adjusted statistically to be equal in the interests of making relative comparisons. That is, the differences in performance among countries reflect differences in student achievement that would be expected on sets of items of equal difficulty.

To give an indication of the difficulty of the reading purpose and process scales, Exhibit A.16 presents the percentage of students responding correctly to each item, averaged across the items for each scale, for each participant. From this exhibit it may be seen that the items making up the literary and informational scales are similar in difficulty—55 percent correct, on average across all participants, for literary reading and 52 percent correct, on average, for informational reading. However, there is a much greater difference in average difficulty between the two scales for the comprehension processes. The average percent correct for the items on the retrieval and straightforward inferencing scale was 64 percent, compared with an average percent correct of just 44 percent for the interpreting, integrating, and evaluating items.

The items making up the interpreting, integrating, and evaluating scale were particularly difficult for students in Kuwait, Morocco, Qatar, and South Africa, where the average percentage of students answering the items correctly ranged from 11 to 14 percent. With average achievement as low as

Exhibit A.16 **Average Percent Correct by PIRLS 2006 Scale**

PIRLS 2006
4th Grade

Countries	Overall	Purposes		Processes	
		Literary	Informational	Retrieval and Straightforward Inferencing	Interpreting, Integrating, and Evaluating
Austria	61 (0.5)	64 (0.6)	59 (0.6)	74 (0.5)	49 (0.6)
Belgium (Flemish)	64 (0.5)	66 (0.6)	62 (0.5)	74 (0.4)	54 (0.5)
Belgium (French)	51 (0.7)	53 (0.7)	49 (0.7)	63 (0.6)	40 (0.7)
Bulgaria	64 (1.1)	65 (1.1)	63 (1.1)	72 (0.9)	57 (1.2)
Canada, Alberta	67 (0.6)	70 (0.6)	64 (0.6)	75 (0.6)	59 (0.6)
Canada, British Columbia	67 (0.6)	69 (0.7)	64 (0.7)	75 (0.6)	59 (0.7)
Canada, Nova Scotia	63 (0.5)	66 (0.6)	60 (0.5)	71 (0.5)	55 (0.6)
Canada, Ontario	66 (0.7)	69 (0.7)	63 (0.8)	73 (0.7)	59 (0.8)
Canada, Quebec	60 (0.7)	62 (0.8)	59 (0.7)	71 (0.6)	50 (0.8)
Chinese Taipei	61 (0.5)	62 (0.5)	60 (0.5)	73 (0.5)	49 (0.5)
Denmark	64 (0.6)	66 (0.6)	61 (0.6)	74 (0.5)	53 (0.6)
England	62 (0.6)	64 (0.7)	60 (0.6)	70 (0.6)	53 (0.6)
France	57 (0.5)	58 (0.5)	57 (0.6)	69 (0.5)	46 (0.6)
Georgia	45 (0.8)	47 (0.9)	42 (0.8)	58 (0.8)	31 (0.8)
Germany	64 (0.5)	67 (0.5)	61 (0.6)	76 (0.5)	53 (0.6)
Hong Kong SAR	69 (0.6)	69 (0.7)	68 (0.6)	77 (0.5)	60 (0.7)
Hungary	65 (0.7)	69 (0.8)	61 (0.7)	73 (0.7)	57 (0.8)
Iceland	54 (0.3)	57 (0.4)	51 (0.4)	67 (0.3)	41 (0.4)
Indonesia	31 (0.7)	29 (0.7)	32 (0.7)	42 (0.8)	20 (0.7)
Iran, Islamic Rep. of	35 (0.6)	37 (0.7)	33 (0.6)	46 (0.7)	23 (0.6)
Israel	56 (0.7)	58 (0.8)	53 (0.7)	65 (0.7)	47 (0.8)
Italy	65 (0.7)	67 (0.8)	63 (0.7)	73 (0.6)	57 (0.8)
Kuwait	22 (0.4)	22 (0.4)	21 (0.5)	30 (0.5)	13 (0.4)
Latvia	63 (0.6)	65 (0.6)	60 (0.6)	71 (0.6)	54 (0.6)
Lithuania	61 (0.4)	65 (0.5)	58 (0.5)	71 (0.4)	52 (0.5)
Luxembourg	66 (0.2)	68 (0.3)	64 (0.3)	78 (0.2)	55 (0.3)
Macedonia, Rep. of	40 (0.8)	39 (0.8)	41 (0.9)	50 (0.8)	29 (0.8)
Moldova, Rep. of	52 (0.8)	52 (0.8)	53 (0.8)	59 (0.8)	45 (0.8)
Morocco	21 (0.7)	20 (0.8)	21 (0.8)	30 (0.9)	11 (0.6)
Netherlands	64 (0.4)	66 (0.4)	62 (0.4)	75 (0.4)	53 (0.4)
New Zealand	60 (0.5)	61 (0.6)	59 (0.5)	68 (0.5)	52 (0.6)
Norway	51 (0.6)	54 (0.6)	48 (0.7)	63 (0.7)	39 (0.6)
Poland	57 (0.6)	59 (0.6)	54 (0.6)	67 (0.6)	47 (0.6)
Qatar	24 (0.2)	24 (0.3)	23 (0.2)	33 (0.2)	14 (0.2)
Romania	50 (1.1)	52 (1.2)	47 (1.1)	60 (1.2)	39 (1.1)
Russian Federation	68 (0.8)	70 (0.8)	66 (0.8)	77 (0.7)	59 (0.9)
Scotland	59 (0.7)	61 (0.8)	57 (0.7)	69 (0.6)	49 (0.8)
Singapore	66 (0.7)	67 (0.8)	66 (0.7)	76 (0.6)	57 (0.8)
Slovak Republic	60 (0.7)	63 (0.8)	57 (0.7)	70 (0.7)	50 (0.7)
Slovenia	57 (0.5)	59 (0.6)	56 (0.6)	68 (0.5)	47 (0.6)
South Africa	21 (0.9)	20 (0.9)	21 (0.8)	28 (0.9)	14 (0.8)
Spain	55 (0.6)	58 (0.7)	52 (0.6)	65 (0.6)	45 (0.7)
Sweden	64 (0.5)	66 (0.6)	62 (0.6)	75 (0.5)	54 (0.6)
Trinidad and Tobago	38 (0.9)	38 (1.0)	38 (0.9)	48 (1.0)	28 (0.9)
United States	62 (0.8)	65 (0.9)	60 (0.8)	70 (0.8)	54 (1.0)
International Avg.	54 (0.1)	55 (0.1)	52 (0.1)	64 (0.1)	44 (0.1)

SOURCE: IEA Progress in International Reading Literacy Study (PIRLS) 2006

( ) Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

this, there is a danger that results may be distorted by a "floor" effect—that the achievement of low-performing students may be overestimated because there are not sufficient items matched to their ability levels. Even though the PIRLS scaling approach is very robust, there is concern that achievement results based on very low average performance may not be sufficiently accurate to report. Based on examinations of the data, PIRLS 2006 used a guideline of not reporting scaled achievement results for countries with an average percent correct of 20 percent or less on the set of items comprising the scale.[22] Accordingly, results on the interpreting, integrating, and evaluating scale were not reported for Kuwait, Morocco, Qatar, and South Africa.

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.[23] As well as sampling error, 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 2 standard errors represents a 95 percent confidence interval for the corresponding population result.

Reporting Student Achievement at the PIRLS 2006 International Benchmarks

To provide richly descriptive information about what performance on the PIRLS reading scale means in terms of the reading skills that students have and the comprehension processes and strategies they can bring to bear, PIRLS identified four points on the scale for use as international benchmarks of

22 For further discussion, see Foy, P., Galia, J., & Li, Isaac. (2007). Scaling the PIRLS 2006 reading assessment data. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2006 technical report*. Chestnut Hill, MA: Boston College.

23 Procedures for computing jackknifed standard errors are presented in Kennedy, A.M. & Trong, K. L. (2007). Reporting PIRLS 2006 student achievement in reading. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2006 technical report*. Chestnut Hill, MA: Boston College.

student achievement. Selected to represent the range of performance shown by students internationally, the advanced benchmark is 625, the high benchmark is 550, the intermediate benchmark is 475, and the low benchmark is 400.

It should be noted that the PIRLS 2006 international benchmarks were established using procedures different from those implemented in 2001. Unlike the PIRLS 2006 benchmarks, which are four fixed points evenly spaced on the scale (625, 550, 475, and 400), the PIRLS 2001 international benchmarks were identified on the basis of student achievement across the participating countries. The most challenging benchmark, the *Top 10% Benchmark*, was defined as the 90th percentile. Corresponding to a scale score of 615, this was the point above which the top 10 percent of students scored, counting across all countries. Next most challenging, the *Upper Quarter Benchmark* was defined as the 75th percentile and corresponded to a scale score of 570, while the *Median Benchmark*, defined as the 50th percentile, or median, corresponded to a scale score of 510, and the *Lower Quarter Benchmark* defined as the 25th percentile, corresponded to a scale score of 435.

Although the PIRLS 2001 approach to establishing benchmarks based on student achievement worked well for the first cycle of PIRLS, for measuring trends across successive cycles of PIRLS it has the disadvantage that, because the benchmarks must be recomputed with each new cycle of the study, benchmarks will change from cycle to cycle depending on the set of countries taking part. For example, if several new low-achieving countries joined a cycle, benchmarks based on percentiles of student achievement could decrease, perhaps giving the erroneous impression that standards had improved. To avoid misinterpretations based on movement in benchmarks between cycles, PIRLS 2006 adopted the fixed benchmark approach, instituted for TIMSS 2003, where the same four scale-score points (625, 550, 475, and 400), will be used as international benchmarks for all future cycles of PIRLS (i.e., in 2011, 2016, and so on).

In order to interpret the PIRLS scale scores and analyze achievement at the international benchmarks, PIRLS 2006 conducted a scale anchoring analysis to describe achievement of students at those four points on the scale.

Scale anchoring is a way of describing students' performance at various 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.[24] In PIRLS 2006, the Reading Development Group worked with the Reading Coordinator, the PIRLS Reading Consultant, and the TIMSS & PIRLS International Study Center to describe student reading at the international benchmarks. The descriptions of the items developed as part of the scale anchoring analysis are found in Appendix E.

24 The scale-anchoring procedure is described fully in Kennedy, A.M. & Trong, K. L. (2007). Reporting PIRLS 2006 student achievement in reading. In M.O. Martin, I.V.S. Mullis, & A.M. Kennedy (Eds.), *PIRLS 2006 technical report*. Chestnut Hill, MA: Boston College.