

# PIRLS

## Appendix A

A

# Appendix A

## Overview of Procedures for the Trends in IEA's Reading Literacy Study

### History

In 1970-71, IEA conducted its first reading study, a study of reading comprehension in 15 countries.<sup>1</sup> Building on the success of this initial venture into reading, IEA embarked on the 1991 Reading Literacy Study,<sup>2</sup> a much more ambitious venture involving extensive testing of two student populations – the grades with most nine- and fourteen-year-olds, respectively. Utilizing a wider range of testing materials than the earlier study, and incorporating detailed questionnaires for students, teachers, and principals, the reading Literacy Study collected data in 1990-91 in 32 countries (27 at the younger age level, and 31 at the older level). PIRLS, the successor to the Reading Literacy Study, was designed not only to provide a state-of-the-art assessment of fourth-grade students' reading literacy achievement in 2001, but also to supply data on a continuous five-year cycle thereafter to monitor progress in reading achievement into the future.

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.

As the PIRLS work on framework development progressed, it became evident that the PIRLS reading assessment would have quite a different emphasis to the Reading Literacy Study, and that it would not be possible to compare results from the two studies directly. As an alternative that would allow countries to measure changes in the reading achievement of their students since 1991, IEA provided PIRLS countries the opportunity to re-administer the 1991 reading literacy test in 2001 – at the same time as the main PIRLS assessment. This study is known as the Trends in IEA's Reading Literacy Study.

### Participants in the Trends in IEA's Reading Literacy Study

Nine of the thirty-five countries participating in the 2001 PIRLS assessment took part also in the trend study in order to examine changes between 1991 and 2001 in student reading performance – as measured by the 1991 reading literacy test (see Exhibit A.1).

### The 1991 Reading Literacy Test

IEA's 1991 reading literacy test was developed through a collaborative process lasting more than two years; and involving the project steering committee, the staff of the international coordinating center at the University of Hamburg, and the national research coordinators from the participating countries.<sup>3</sup> The specifications for the test comprised three major domains, corresponding to three types of text presumed to cover the main varieties of reading materials encountered by young children in most countries: narrative texts, expository texts, and documents.

- Narrative texts include continuous textual materials in which the writer's aim is to tell a story – whether factual or fictional. Narrative texts normally are designed to entertain or involve the reader emotionally; are written in the past tense; and usually have people or animals as their main theme.
- Expository texts are designed to describe or explain something; they may be written in the present or past tense; and the style is typically formal and impersonal – highlighting such features as definitions, causes, classifications, functions, contrasts, and examples.

<sup>3</sup> The description of the reading literacy test development process provided in this report was abstracted from Elley, W.B. (1995). The measurement of reading literacy: How the international tests of literacy were developed. In R.M. Wolf (Ed.), *The IEA reading literacy study: Technical report*. The Hague: International Association for the Evaluation of Educational Achievement.

**Exhibit A.1: Countries Participating in the Trends in IEA's Reading Literacy Study**ISC  
RLS Trend  
1991–2001

Country	Country's Name for Grade Tested	Years of Formal Schooling	Average Age of Students Tested in 2001
Greece	4	4	10.0
Hungary	3	3	9.7
Iceland	4	4	9.8
Italy	4	4	9.9
New Zealand	Year 5 <sup>1</sup>	4	10.0
Singapore	Primary 3	3	9.1
Slovenia	3	3	9.8
Sweden	3	3	9.8
United States	4	4	10.0

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

**Exhibit A.2: Blueprint of Items by Domain for the 1991 Reading Literacy Test**

Domain	Verbatim	Paraphrase	Inference	Locate Information	Locate and Process	Total
Narrative text	1	11	10	–	–	22
Expository text	7	9	5	–	–	21
Documents	–	–	–	11	12	23
<b>Total Items</b>	<b>8</b>	<b>20</b>	<b>15</b>	<b>11</b>	<b>12</b>	<b>66</b>

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

<sup>1</sup> The official nomenclature used in New Zealand since 1996 refers to students' years of schooling rather than a class/grade level. Year 5 students were at a class level equivalent to Grade 4.

- Documents refer to such things as forms, charts, labels, graphs, recipes, labels, maps, directories, and sets of instructions. Students usually are required to skim the text to identify its structure, and use that to locate required information.

Exhibit A.2 shows the blueprint for the test, with items classified by text type, and by the various skills or activities students were assumed to use in responding to each item. The narrative text had four passages with 22 items; the expository text had five passages with 21 items; and there were six documents with 23 items. Of the 66 items, eight required a verbatim response (i.e., the answer resided in the text in much the same wording as in the question). Twenty items required students to paraphrase or recognize the answer in the text in different wording from that of the question; 15 items required students to go beyond the information given and make an inference in arriving at the correct answer. In the documents, 11 items required students merely to locate a fact or figure, while a further 12 asked them to locate and process (count, compare, or infer). Of the items in total, four required the student to write a word or phrase; two required an extended written response;<sup>4</sup> and the remainder (60) were in multiple-choice format (with four options for each item).

The test consisted of two student booklets, and was administered in two sessions of 35 and 40 minutes, respectively. The beginning of the first booklet contained a short word recognition test (40 items to be completed in 90 seconds). In keeping with 1991 data-collection procedures, the word recognition test was administered also in 2001, but the results were not included in the analysis of the trend study data.

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. In selecting the passages for the reading literacy test, every effort was made to minimize cultural bias. Potential stimulus passages and items 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. Everything possible was done to ensure that the items did not exhibit bias towards or against particular countries. Draft passages and

4 The two extended response items were administered in 1991 but not scored or included in students' results. The same procedure was followed in 2001.

items were subjected to full-scale field testing before the instruments for the main data collection were finalized.

### Student Questionnaire

The student questionnaire<sup>5</sup> asked students about their home circumstances; it included questions about their possessions in the home, home literacy resources, home literacy interactions, out-of-school activities, and beliefs about reading. Students also were asked about their voluntary reading habits, and about their in-school reading habits.

### Translation of Tests and Questionnaires

The reading literacy instruments were prepared in English, then translated by national centers into the local language of instruction. Countries were provided with explicit guidelines for translation and cultural adaptation, which required independent translations by two expert translators familiar with age-appropriate linguistic demands. An extensive series of statistical checks were conducted after the testing, to detect items not performing comparably across countries or over time.

### Sample Implementation and Participation Rates

IEA's 1991 Reading Literacy Study targeted primary/elementary-level students enrolled in the grade containing the largest proportion of 9-year-old students at the time of testing – generally the third or fourth grade in each country.<sup>6</sup> To maintain comparability, the same population was targeted by the trend study for testing in 2001.<sup>7</sup> Exhibit A.3 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 international comparative studies such as PIRLS or the trend study. The accuracy of the survey results depends on the quality of the sampling information available when planning the sample, and on the care with which the sampling activities are conducted. The sampling for the trend study was conducted in parallel with the PIRLS 2001 sampling. NRCs worked on all phases

5 The 1991 Reading Literacy Study included extensive questionnaires for students, teachers, and school principals. Only the student questionnaire was administered in the 2001 data collection.

6 Ross, K.N. (1995). Sample design procedures for the international study of reading literacy. In R.M. Wolf (Ed.), *The IEA reading literacy study: technical report*. The Hague: International Association for the Evaluation of Educational Achievement.

7 See Foy, P., & 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.

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 all aspects of sampling for the trend study – including the national sampling plans, sampling data, sampling frames, and sample selection. 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 PIRLS 2001 sampling design was a two-stage stratified cluster sample, with a sample of schools as the first stage and a sample from the classrooms from the target grade in those schools as the second stage. For efficiency of sampling, the trend study adopted the same basic design; and it worked from the same sample of schools. For PIRLS, most countries sampled 150 schools and one intact classroom from each school, although some countries selected larger samples.<sup>8</sup> The school sample for the trend study consisted of half the schools (every other school) sampled for the PIRLS data collection. From each of these schools, an additional classroom was sampled from the target grade for use in the trend data collection.

Exhibits A.4 and A.5 present achieved sample sizes for schools and students, respectively. Exhibit A.6 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 all linked to the students' achievement data.

8 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.3: Population Coverage and Exclusions – Trends in IEA's Reading Literacy Study**ISC  
RLS Trend  
1991–2001

Country	International Desired Population Coverage	National Desired Population		
		School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Greece	100%	2.0%	4.0%	6.0%
Hungary	100%	1.8%	0.0%	1.8%
Iceland	100%	1.8%	2.0%	3.8%
Italy	100%	0.0%	3.4%	3.4%
New Zealand <sup>1</sup>	100%	1.6%	1.3%	2.9%
Singapore	100%	1.3%	0.0%	1.3%
Slovenia	100%	0.0%	0.9%	0.9%
Sweden	100%	2.5%	2.2%	4.7%
United States	100%	0.6%	3.9%	4.5%

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

**Exhibit A.4: School Participation Rates and Sample Sizes – Trends in IEA's Reading Literacy Study**ISC  
RLS Trend  
1991–2001

Country	School Participation Before Replacement (Weighted Percentage)	School Participation After Replacement (Weighted Percentage)	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample That Participated	Total Number of Schools That Participated
Greece	73%	79%	85	85	63	68
Hungary	98%	98%	220	220	216	216
Iceland	93%	93%	70	70	65	65
Italy	89%	100%	92	92	81	92
New Zealand	90%	98%	75	75	67	73
Singapore	100%	100%	98	98	98	98
Slovenia	100%	100%	75	75	75	75
Sweden	96%	100%	150	150	142	148
United States	58%	85%	100	100	54	85

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

1 The Maori school stratum was not part of the study.

**Exhibit A.5: Student Participation Rates and Sample Sizes – Trends in IEA's Reading Literacy Study**
**ISC** RLS Trend  
1991–2001

Country	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
Greece	97%	1195	0	47	1148	39	1109
Hungary	97%	4859	20	0	4839	132	4707
Iceland	86%	2137	14	44	2079	282	1797
Italy	97%	1697	6	56	1635	45	1590
New Zealand <sup>1</sup>	95%	1308	43	19	1246	58	1188
Singapore	98%	3729	46	0	3683	82	3601
Slovenia	95%	1577	0	2	1575	73	1502
Sweden	96%	5706	33	118	5555	194	5361
United States	95%	1980	20	40	1920	94	1826

SOURCE: Trends in IEA's Reading Literacy Study 1991–2001

**Exhibit A.6: School and Student Participation Rates (Weighted) – Trends in IEA's Reading Literacy Study**
**ISC** RLS Trend  
1991–2001

Country	School Participation Before Replacement	School Participation After Replacement	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Greece	73%	79%	97%	70%	77%
Hungary	98%	98%	97%	96%	96%
Iceland	93%	93%	87%	80%	80%
Italy	89%	100%	97%	86%	97%
New Zealand <sup>1</sup>	90%	98%	95%	85%	93%
Singapore	100%	100%	98%	98%	98%
Slovenia	100%	100%	95%	95%	95%
Sweden	96%	100%	97%	93%	97%
United States	58%	85%	95%	55%	81%

SOURCE: Trends in IEA's Reading Literacy Study 1991–2001

<sup>1</sup> The Maori school stratum was not part of the study.

## 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 the scripts used to regulate directions and timing, rules for answering students' questions, and steps ensuring that identification on the test booklets and questionnaires corresponded to the information on the forms used to track students. Countries also were expected to conduct quality control visits to a sample of the trend study schools, as part of their national quality control program for PIRLS.

## Test Reliability

Exhibit A.7 displays the reading test Cronbach's alpha reliability coefficient for the reading literacy test in 1991 and 2001 for each country. Reliabilities were acceptably high in all countries, ranging from 0.91 to 0.95.

## Data Processing

To ensure the availability of comparable, high-quality data for the trend study analysis, the study implemented rigorous quality control in creating the international database.<sup>9</sup> Manuals and software were prepared 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 inclusion in 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 achievement and questionnaire data files.

9 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.

**Exhibit A.7: Cronbach's Alpha Reliability Coefficient – Trends in IEA's Reading Literacy Study**

ISC

RLS Trend  
1991–2001

Countries	Reliability Coefficient	
	1991	2001
Greece	0.92	0.92
Hungary	0.93	0.93
Iceland	0.94	0.92
Italy	0.93	0.92
New Zealand	0.94	0.94
Singapore	0.91	0.93
Slovenia	0.93	0.92
Sweden	0.95	0.94
United States	0.91	0.92
International Median	0.93	0.92

SOURCE: Trends in IEA's Reading Literacy Study 1991–2001

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 were given several opportunities to review the data for their countries. In conjunction with the IEA Data Processing Center, the International Study Center reviewed item statistics for each cognitive item in each country to identify poorly performing items. In general, the items exhibited very good psychometric properties in all countries, although one or two items in a few countries had properties in the 2001 data different from in 1991, and were, therefore, eliminated from the trend analysis.<sup>10</sup>

### IRT Scaling and Data Analysis

The general approach to reporting the achievement data from the PIRLS and the trend study was based primarily on item response theory (IRT) scaling methods.<sup>11</sup> Student reading achievement in PIRLS was summarized using a family of 2-parameter and 3-parameter IRT models 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 in a way that takes into account the difficulty and discriminating power of each item. The 3-parameter IRT methodology used with PIRLS also was applied in scaling the trend study data, placing the data from both 1991 and 2001 on the same scale so that changes in students' average reading achievement over the ten-year period could be described accurately. The PIRLS methodology was used partly for consistency with the PIRLS approach, but mainly because it was judged to provide the most accurate estimates of change in student reading achievement.

By combining the data from 1991 and 2001 in a single analysis, the IRT approach provides a common scale on which performance can be compared over time, as well as 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. To provide a basis for comparison, the average of the scale scores

10 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. Appendix C of the technical report contains a list of items eliminated from the scaling.

11 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.

for the 2001 data, across countries, was set to 500, and the standard deviation to 100. Since the countries varied in size, each country was weighted to contribute equally to the mean and standard deviation of the scale. Results from 1991 were then placed on this scale also, so that changes in student performance between 1991 and 2001 would be readily apparent. Four separate scales were constructed for the trend study: one for each of the narrative, expository, and documents domains, and one for reading achievement overall.

To allow more accurate estimation of summary statistics for student subpopulations, the PIRLS and trend study 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 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.

### 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.<sup>12</sup> 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 reflecting 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.

12 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.

