PIRLS

A e di A


A

Appendix A

1 T[illegible], R.L. (1973). R[illegible]. [illegible] : A[illegible]. *International studies in evaluation: Vol. 3*. S[illegible] : A[illegible] & W[illegible].

2 E[illegible], W.B. (E[illegible].). (1994). *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems*. O[illegible], E[illegible] : E[illegible] S[illegible] L[illegible].

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.

E 64 2a.1: C 1 Pa a 1 T IEA' R1a L 1

- 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.,

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

Student Questionnaire

The student questionnaire[5] 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 Reading Literacy Instruments

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.[6] To maintain comparability, the same population was targeted by the trend study for testing in 2001.[7] 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 ... 2001 ...

6 ..., K.N. (1995). ... *The IEA reading literacy study: technical report*. ...

7 ..., P., & ..., M. (2003). PIRLS ... M.O. M..., I.V.S. M..., & A.M. K... (Eds.), *PIRLS 2001 technical report*. ...

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.[8] 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 F . ,n , J , M. (2003). PIRLS . . n . I M.O. M , I.V.S. M. , & A.M. K (E .), *PIRLS 2001 technical report*. Cn . H , MA: B C . .

E : A.3: P a C a E T IEA' Ra L a S

ISC RLS Trend 1991–2001

Country	International Desired Population Coverage	National Desired Population		
		School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
G eece	100%	2.0%	4.0%	6.0%
H a	100%	1.8%	0.0%	1.8%
Ice a d	100%	1.8%	2.0%	3.8%
I a	100%	0.0%	3.4%	3.4%
Ne Zea a d [1]	100%	1.6%	1.3%	2.9%
S a e	100%	1.3%	0.0%	1.3%
S e a	100%	0.0%	0.9%	0.9%
S ede	100%	2.5%	2.2%	4.7%
U ed S a e	100%	0.6%	3.9%	4.5%

SOURCE: IEA P (PIRLS) 2001. S L R

E : A.4: S Pa a Ra a Sa S T IEA' Ra L a S

ISC

SOURCE: IEA P (PIRLS) 2001. S L R

1 To M

E : A.7: C :a ' A a R a: C ff T IEA' R a
L a S

ISC RLS Trend 1991–2001

Countries	Reliability Coefficient	
	1991	2001
G eece	0.92	0.92
H a	0.93	0.93
Ice a d	0.94	0.92
I a	0.93	0.92
Ne Zea a d	0.94	0.94
S a e	0.91	0.93
S e a	0.93	0.92
S ede	0.95	0.94
U ed S a e	0.91	0.92
I e a a Med a	0.93	0.92

SOURCE: T IEA' R L S 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.[10]

IRT S a a Da a A a

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.[11] 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 analgin sthe r079.8(y be0l9dter IR)49.8(y witl
in st280069 Tc0 be0l9d stud.8(y 0011(w countrva fr)1ar9.8(y)10ng), to pr)19.7(o)25.1(vinels f)19.ma(

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-