12.1   Overview

The *PIRLS 2001 International Report* (Mullis, Martin, Gonzalez, and Kennedy, 2003) summarizes fourth-grade students' reading achievement in each country. This chapter provides information about how important statistics in the report were computed, including their standard errors; describes how international benchmarks of achievement were established to facilitate reporting achievement, outlines the scale-anchoring procedure followed to describe performance at these benchmarks; and describes briefly the reporting of the information collected by questionnaire from the students and their parents, teachers, and school principals.

12.2   Estimation of Sampling and Imputation Variance

To obtain estimates of students' reading proficiency that were both accurate and cost-effective, PIRLS 2001 made extensive use of probability sampling techniques to sample students from national fourth-grade student populations, and of matrix sampling methods to target individual students with a subset of the entire set of assessment materials. Statistics computed from these student samples were used to estimate population parameters. This approach made an efficient use of resources, in particular keeping student response burden to a minimum, but at a cost of some variance or uncertainty in the statistics. To quantify this uncertainty, each statistic in the *PIRLS 2001 International Report*

and in the trend study report, in *Trends in Children's Reading Literacy Achievement 1991–2001* (Martin, Mullis, Gonzalez, and Kennedy, 2003) is accompanied by an estimate of its standard error. These standard errors incorporate components reflecting the uncertainty due to generalizing from student samples to the entire fourth-grade student population (sampling variance), and to inferring students' performance on the entire assessment from their performance on the subset of items that they took (imputation variance).

12.2.1 Estimating Sampling Variance

The PIRLS 2001 sampling design applied Tw[(9 g)24.9(enel design a)19.6(ppli77on the54.btand7(u -1.25lti)19m s cTj9m st7(gm st8(e)1.25snts r). -1.25 495.9999 Tm-09[(The9 3*9905 745.323112.2.1)-820.8(Es4.557270025 Tw[((Ma

	2001	1 1
ꓬ ꓧ ꓧ	69	. .
ꓶ ꓶ	60	. .
ꓶ ꓹ ꓵ	75	. .
ꓧ	75	. .
	74	. .
ꓸ	75	. .
	71	. .
ꓧ ꓧ	66	. .
ꓲꓹ ꓧ	73	. .
ꓲ ꓹ ꓧ	75	. .
ꓲ ꓹ		

96 ꓹ 0. 2 .93 0 . (.)6

ꓶ ꓹ ꓹ ꓶ 4ꓲ   . 6-5 3412.96 ꓧ   ꓢ   5266

Construction of Sampling Zones

To apply the JRR technique used in PIRLS 2001, the sampled schools are paired and assigned to a series of groups known as sampling zones. This was done at Statistics Canada by working through the list of sampled schools in the order in which they were selected and assigning the first and second schools to the first sampling zone, the third and fourth schools to the second zone, and so on. In total, 75 zones were used, allowing for 150 schools per country. When more than 75 zones were constructed, they were collapsed to keep the total number to 75.

Sampling zones were constructed within design domains, or explicit strata. Where there was an odd number of schools in an explicit stratum, either by design or because of school nonresponse, the students in the remaining school were randomly divided to make up two "quasi" schools for the purposes of calculating the jackknife standard error. Each zone then consisted of a pair of schools or "quasi" schools. Exhibit 12.1 shows the number of sampling zones used in each country.

Computing Sampling Variance Using the JRR Method

The JRR algorithm used in PIRLS 2001 assumes that there are H sampling zones within each country, each containing two sampled schools selected independently. To compute a statistic t from the sample for a country, the formula for the JRR variance estimate of the statistic t is then given by the following equation:

where H is the number of pairs in the sample for the country. The term $t(S)$ corresponds to the statistic for the whole sample (computed with any specific weights that may have been used to compensate for the unequal probability of selection of the different elements in the sample or any other post-stratification weight). The element $t(J_h$

weights were not permanent variables, but were created temporarily by the sampling variance estimation program as a useful computing device.

To create replicate weights, each sampled student was first assigned a vector of 75 weights where h takes values from 1 to 75. The value of is the overall sampling weight, which is simply the product of the final school weight, the appropriate final classroom weight, and the appropriate final student weight, as described in Chapter 9.

The replicate weights for a single case were then computed as:

where the variable k_h for an individual i takes the value $k_{hi} = 2^*$

The general procedure for estimating the imputation variance using plausible values is the following (Mislevy et al., 1992). First compute the statistic (*t*) for each set of plausible values (*M*). The statistics t_m, where $m = 1, 2, \ldots, 5$, can be anything estimable from the data, such as a mean, the difference between means, percentiles, and so forth. Each of these statistics will be called t_m.

Once the statistics are computed, the imputation variance is then computed as:

$$Var_{imp} = \left(1 + \frac{1}{M}\right) Var(t_m)$$

where *M* is the number of plausible values used in the calculation, and $Var(t_m)$ is the variance of the estimates computed using each plausible value.

12.2.3 Combining Sampling and Imputation Variance

The standard errors of the reading proficiency statistics reported by PIRLS include both sampling and imputation variance components. The standard errors were computed using the following formula:[3]

$$Var \cdot (t_{pv}) = Var_{jrr}(t_1) + Var_{imp}$$

where $Var_{jrr}(t_1)$ is the sampling variance for the first plausible value and Var_{imp} in the imputation variance. The forthcoming User Guide for the PIRLS 2001 International Database contains programs in SAS and SPSS that compute each of these variance components for the PIRLS 2001 data.

Exhibits 12.2 through 12.4 show basic summary statistics for reading achievement in the PIRLS 2001 assessment, for reading overall, as well as for reading for literary

	3300	420	96	5.8	5.9
	2909	327	106	4.6	4.7
	3460	550	83	3.8	3.8
	8253	544	72	2.3	2.4

	3300	419	97	5.8	5.8
	2909	330	103	4.7	4.9
	3460	550	86	3.9	3.9
	8253	545	75	2.5	2.6
	5131	425	79	4.2	4.2
	3001	498	80	2.5	2.5
	3022	535	63	2.2	2.3
	3156	559	94	3.6	3.9
	3538	518	71	2.5	2.6
	7633	537	66	1.9	1.9
	2494	528	74	3.3	3.3
	5050	518	66	3.0	3.1
	4666	548	65	2.0	2.0
	3676	520	69	1.13.351.8788 -1.697	

Exhibit 12.4: Summary Statistics and Standard Errors for PIRLS 2001 Reading to Acquire and
Use Information

	3300	422	99	5.4	5.4
	2909	332	109	4.9	4.9
	3460	551	81	3.4	3.6
	8253	541	71	2.3	2.4
	5131	424	83	4.2	4.3
	3001	490	83	2.9	3.0
	3022	536	68	2.5	2.7
	3156	546	82	3.4	3.6
	3538	533	71	2.4	2.5
	7633	538	68	1.8	1.9
	2494	521	75	3.7	3.7
	5050	537	59	2.8	2.9
	4666	537	68	2.2	2.2
	3676	504	84	1.2	1.5
	7430	408	97	4.6	4.6
	3973	507	93	2.8	2.9
	3502	536	69	2.3	2.4
	7126	403	97	4.5	4.5
	3019	547	64	2.2	2.3
	2567	540	64	2.5	2.7
	3711	445	108	5.1	5.2
	3533	505	81	4.5	4.7
	3153	358	125	10.8	10.9
	4112	553	58	2.4	2.6
	2488	525	89	3.5	3.8
	3459	492	81	2.8	2.8
	3625	512	90	4.6	4.6
	4093	531	68	4.3	4.3
	2717	527	82	3.4	3.6
	7002	527	83	4.8	4.8
	3807	522	71	2.7	2.7
	2952	503	75	1.8	1.9
	6044	559	68	2.1	2.2
	5125	452	90	3.8	3.8
	3763	533	79	3.5	3.7

Exhibit 12.5: Summary Statistics and Standard Errors for IEA's Trends in Reading Literacy Study – Overall Reading Achievement

	2001				
				.	
- ﹐	1109	507	91	5.8	5.9
ﻪ ﹐	4707	475	97	3.8	3.9
ﻪ	1797	513	94	3.3	3.5
	1590	513	92	4.4	4.4
﹀ ﻪ	1188	502	111	5.2	5.3
�.ﻪ ﹐	3601	489	106	7.9	8.0
ﺺ . ﻪ	1502	493	91	3.7	3.7
ﺺ ﻪ	5361	498	115	3.8	3.9
_ ﻪ ﺺ	1826	511	94	6.3	6.3

	1 1				
				.	
- ﹐	3516	466	96	4.5	4.5
ﻪ ﹐	3009	459	93	3.9	4.0
ﻪ	3961	486	104	1.4	1.5
	2221	500	101	5.3	5.4
﹀ ﻪ	3016	498	110	4.1	4.1
ﺺ.ﻪ ﹐	7326	481	88	3.5	3.6
ﺺ . ﻪ	3297	458	96	3.2	3.2
ﺺ ﻪ	4301	513	116	4.2	4.2
_ ﻪ ﺺ	6433	521	90	3.2	3.2

Exhibit 12.6: Summary Statistics and Standard Errors for IEA's Trends in Reading Literacy Study – Reading Narrative Texts

	2001				
- ,	1109	513	88	4.7	4.8
,ђ ,	4707	479	85	3.1	3.1
,ђ	1797	524	100	3.2	3.3
	1590	517	88	3.9	4.1
ٔ ,ђ	1188	496	114	5.3	5.3
Ş,ђ ,	3601	487	113	8.6	8.6
Ş . ,ђ	1502	490	88	3.4	3.7
Ş ,ђ	5361	496	104	3.2	3.6
_ ,ђ Ş	1826	498	105	6.6	6.8

	1 1				
- ,	3516	479	87	3.6	3.7
,ђ ,	3009	467	81	3.1	3.2
,ђ	3961	493	98	1.4	1.6
	2221	507	91	4.6	4.7
ٔ ,ђ	3016	500	111	4.2	4.3
Ş,ђ ,	7326	486	94	3.5	3.5
Ş . ,ђ	3297	465	90	2.9	3.0
Ş ,ђ	4301	513	100	3.3	3.4
_ ,ђ Ş	6433	518	101	3.2	3.3

- ⟩	1109	509	91	5.1	5.2
⸂ ⟩	4707	464	111	4.3	4.4
⸂	1797	502	97	3.1	3.3
	1590	513	99	4.4	4.5
⟍ ⸂	1188	510	101	5.2	5.3
�?⸂ ⟩	3601	495	91	6.5	6.6
⸉ . ⸂	1502	489	92	3.1	3.3
⸉ ⸂	5361	496	121		

	2001				
‒ ›	1109	490	92	5.1	5.2
‎ʰ ›	4707	486	102	3.7	3.7
‎ʰ	1797	506	89	3.2	3.4
	1590	499	93	4.4	4.5
‎\ ʰ	1188	506	113	4.9	5.2
ʂʰ ›	3601	484	96	6.8	6.8
ʂ . ʰ	1502	502	92	3.6	3.8
ʂ ʰ	5361	506	122	3.9	4.4
‒ ʰ ʂ	1826	520	90	5.9	6.1

12.3 Reporting Student Achievement in Reading

As described in earlier chapters, PIRLS made extensive use of imputed proficiency scores to report student achievement in reading, for each of the two reading purposes – reading for literary experience and to acquire and use information – and for reading overall. This section describes the procedures followed in computing the principal statistics used to summarize achievement in the *PIRLS 2001 International Report* (Mullis, Martin, Gonzalez, & Kennedy, 2003), including country means based on plausible values, international benchmarks of achievement, gender differences, and performance on example items. It also presents means and standard errors for the nine

then averaging across countries. These five
estimates were then averaged to derive the
international average presented in the
PIRLS reports, as shown below:

where

is the international mean for plau-

Exhibit 12.9: Means and Standard Errors for International Comparisons – PIRLS 2001

	419.527	5.935	419.187	5.792	422.417	5.448
	326.829	4.697	329.596	4.853	332.175	4.947
	550.498	3.847	549.542	3.866	551.310	3.573
	544.146	2.377	544.567	2.609	541.300	2.449
	422.428	4.447	425.326	4.248	423.629	4.283
	493.976	2.982	498.129	2.532	489.898	2.970
	536.883	2.321	535.287	2.335	536.399	2.680
	552.878	3.394	559.177	3.883	545.624	3.557
	525.170	2.367	518.149	2.642	533.133	2.537
	539.090	1.935	536.515	1.942	538.181	1.949
	524.167	3.487	527.640	3.345	520.986	3.707
	527.871	3.079	517.553	3.063	537.238	2.933
	543.226	2.199	548.462	2.031	537.273	2.199
	512.417	1.199	520.071	1.307	504.089	1.467
	413.833	4.182	420.843	4.470	408.398	4.642
	508.939	2.835	510.049	2.598	506.763	2.880
	540.729	2.352	543.101	2.697	536.155	2.357
	396.471	4.295	393.803	3.824	403.247	4.542
	544.607	2.284	537.206	2.177	546.946	2.345
	543.387	2.589	545.518	3.086	539.544	2.677
	441.586	4.610	441.477	4.457	445.321	5.200
	491.743	3.967	479.938	3.703	504.888	4.688
	349.511	9.650	347.148	8.352	358.014	10.855
	554.209	2.497	552.285	2.494	552.834	2.621
	528.824	3.563	531.368	3.880	524.857	3.825
	499.179	2.922	505.703	2.750	492.133	2.836
	511.710	4.589	511.822	4.727	512.424	4.598
	527.933	4.432	523.490	3.870	531.450	4.323
	528.176	3.601	529.097	3.543	527.033	3.605
	527.948	5.156	528.483	5.565	527.356	4.803
	518.087	2.846	512.119	2.581	522.135	2.709
	501.518	1.966	499.358	1.816	503.123	1.924
	561.014	2.218	559.403	2.383	558.605	2.212
	449.354	3.537	448.186	3.377	451.811	3.797
	542.149	3.817	550.408	3.812	533.325	3.655

included in the adjustment, leading to apparently conflicting results from comparisons using different numbers of countries.

12.3.2 Comparing Achievement with the International Mean

Many of the data exhibits in the PIRLS 2001 international reports show countries' mean achievement compared with the international mean, together with a test of the statistical significance of the difference between the two. These significance tests are based on the standard errors of the national and international means.

When comparing each country's mean with the international average, PIRLS took into account the fact that the country contributed to the international standard error. To correct for this contribution, PIRLS adjusted the standard error of the difference. The sampling component of the standard error of the difference for country j was:

$$S_{s\_dif\_j} = \frac{\sqrt{\left((N-1)^2 - 1\right)se_j^2 + \sum_{k=1}^{K} se_k^2}}{K}$$

where

$se_{s\_dif\_j}$ is the standard error of the difference due to sampling when country j is compared to the international mean

K is the number of countries

se_j^2 is the sampling standard error for country j

is the sampling standard error for country k

The imputation component of the standard error was computed by taking the square root of the imputation variance calculated as follows

where d_l is the difference between the international mean and the country mean for plausible value l.

Finally, the standard error of the difference was calculated as:

$$se \qquad \sqrt{se \qquad se}$$

12.3.3 International Benchmarks of Reading Achievement

In order to provide information about the range of fourth-grade student reading achievement, PIRLS identified four points on the overall reading scale for use as international benchmarks, and reported the percentage of students reaching these benchmarks in each country. These four points correspond to the 90th, 75th, 50th, and 25th international percentiles of students achievement. The Top 10 percent Benchmark was defined as the 90th percentile on the PIRLS reading scale, computed across all students in all participating countries, with countries weighted in proportion to the size of their fourth-grade population. This point on the scale is the point above which the top 10 percent of students in the 2001 PIRLS assessment

scored. If student reading achievement was distributed in the same way across all countries, approximately 10 percent of students within each country would be above the

12.3.4 Gender Differences

PIRLS reported gender differences in student achievement in reading overall, as well as in the two reading purposes. Gender differences were presented in an exhibit showing the percentages of males and females and their mean reading achievement in each country, together with an indication of whether the male-female difference in reading achievement was statistically significant. Because in most countries males and females attend the same schools, the samples of males and females cannot be treated as independent for the purpose of statistical significance testing. Accordingly, PIRLS used a jackknife procedure applicable to correlated samples for estimating the standard error of the male-female difference. This involved computing the average difference between boys and girls once for each of the 75 replicate samples, and five more times, once for each plausible value, as described earlier in this chapter.

12.3.5 Reporting Student Performance on Individual Items

To portray student achievement as fully as possible, the PIRLS 2001 international report presents many examples of the items used in the assessment, together with the percentage of students in each country responding correctly to or earning partial

was not included in the student's booklet or had been mistranslated or misprinted. The percent correct for an item (P_j) was computed as:

where c_j, w_j, i_j, r_j and n_j are the weighted counts of the correct, wrong, invalid, not reached, and not interpretable responses to item j, respectively.

12.3.6 Trends in Achievement on the IEA

For the 75th percentile (the Upper Quarter Benchmark), an item anchored if:

- At least 65 percent of students scoring in the range answered the item correctly and

- Less than 50 percent of students at the 50th percentile answered the item correctly

For the 90th percentile (the Top 10% Benchmark), an item anchored if:

- At least 65 percent of students scoring in the range answered the item correctly and

- Less than 50 percent of students at the 75th percentile answered the item correctly

To supplement the pool of anchor items, items that met a slightly less stringent set of criteria were also identified. The criteria to identify items that "almost anchored" were the following:

For the 25th percentile, an item almost anchored if:

- At least 60 percent of students scoring in the range answered the item correctly

- Because the 25th percentile is the lowest point, items were not identified in terms of performance at a lower point

For the 50th percentile, an item almost anchored if:

- At least 60 percent of students scoring in the range answered the item correctly and

- Less than 50 percent of students at the 25th percentile answered the item correctly

For the 75th percentile, an item almost anchored if:

- At least 60 percent of students scoring in the range answered the item correctly and

- Less than 50 percent of students at the 50th percentile answered the item correctly

For the 90th percentile, an item almost anchored if:

- At least 60 percent of students scoring in the range answered the item correctly and

- Less than 50 percent of students at the 75th percentile answered the item correctly

To further supplement the pool of items, items that met only the criterion that at least 60 percent of the students answered correctly (regardless of the performance of

Computing the Item Percent Correct at Each Level

The percentage of students scoring in the range around each anchor point that answered the item correctly was computed. To that end, students were weighted to contribute proportionally to the size of the student population in a country. About half of the PIRLS 2001 items are scored dichotomously. For these items, the percentage of students at each anchor point who answered each item correctly was computed. Some of the open-ended items, however, are scored on a partial-credit basis (one, two, or three points); these were transformed into a series of dichotomously scored items, as follows. Consider an item that was scored 0, 1, or 2. Two variables were created:

- $v_1 = 1$ if the student receives a 1, or 2, and 0 otherwise

- $v_2 = 1$ if the student receives a 2 and 0 otherwise.

The percent of students receiving a 1 on v_1 and the percentage of those receiving a 1 on v_2 were computed. This yielded the percent of students receiving at least one point, and full credit.

Identifying Anchor Items

For the PIRLS 2001 reading scale, the criteria described above were applied to identify the items that anchored, almost anchored, and met only the 60 to 65 percent criterion. Exhibits 12.14 and 12.15 present the num-

Exhibit 12.15: Number of Constructed-Response Point Values Anchoring at Each Anchor Level

25 , ,h,	15
50 , ,h,	31
75 , ,h,	17
90 , ,h,	11
, , , 90	13

Exhibit 12.16 presents, by reading purpose, the number of items that met the anchoring criteria discussed above, at each international percentile, and the number of items that were too difficult for the 90th percentile.

12.4.2 Review of Anchor Items Development of Anchor Level Descriptions

Having identified the items that anchored at each of the international benchmarks, the next step was to have the items reviewed by reading experts to develop descriptions of the level of reading comprehension the items demand. In view of their extensive experience in reading and their thorough knowledge of the PIRLS frameworks and achievement tests, the PIRLS Reading Development Group (RDG)

was asked to perform this task. In preparation for the review by the RDG, the items were organized in binders grouped by benchmark anchor point and within anchor point by reading purpose, each binder having four sections, corresponding to the four anchor points. Within each section, the items were sorted by reading purpose and then by the anchoring criteria they met – items that anchored, followed by items that almost anchored, followed by items that met only the 60 to 65 percent criteria. The following information was included for each item: its PIRLS 2001 reading purpose and reading process categories; its answer key; percent correct at each anchor point; and overall international percent correct. For constructed-response items, the scoring guides were included.

The PIRLS International Study Center convened the RDG for a three-day meeting. The assignment consisted of three tasks: (1) work through each item in each binder and arrive at a short description of the knowledge, understanding, and/or skills demonstrated by students answering the item correctly; (2) based on the items that anchored, almost anchored, and met only the 60–65 percent criterion, draft a description of the level of comprehension demonstrated by students at each of the four

Exhibit 12.16: Number of Point Values Anchoring* at Each Anchor Level, by Reading Purpose

	25	50	5	0	0	
,h ,——,h, ,	12	19	20	7	9	67
', , , ,	17	25	13	7	4	66

Although the vast majority of the PIRLS classes were taught by a single teacher, in Sweden each class had two teachers, each of which completed a teacher questionnaire. For reporting in these cases, the student's sampling weight was divided between the teachers, so that the student's contribution to student population estimates thus remained constant regardless of the number of teachers. This was consistent with the policy of reporting attributes of teachers and their classrooms in terms of the percentages of students taught by teachers with these attributes.

12.5.3   Reporting Parents' Questionnaire Data

The PIRLS *Learning to Read Survey* was completed by the parents or primary caregivers of the students participating in the study. Like the teacher questionnaire, the data from the parents' questionnaire were linked to the student, who was always the unit of analysis, even when information from the parents' questionnaires was being reported. That is, the data presented are the percentages of students whose parents reported various characteristics or instructional strategies.

12.5.4   Reporting School Questionnaire Data

The principals of the selected schools in PIRLS completed questionnaires on the school contexts in which the learning and teaching of reading occur. Although schools constituted the first stage of sampling, the PIRLS school sample was

designed to optimize the student sample, not to provide an optimal sample of schools.[6] Therefore, like the teacher data, the school-level data were reported using the student as the unit of analysis to describe the school contexts for the representative samples of students. In general, the exhibits based on the school data present percentages of students in schools with different characteristics for each country and for the international average.

12.5.5   Reporting Response Rates for

of the required background variables were missing. However, for the 10 indices described earlier in this chapter, cases were coded as missing only if there was no response for more the one-third of the questions used to compute the index; index values were be computed if there were valid data for at least two-thirds of the required variables.

The tables in the PIRLS international reports contain special notations on response rates for the background variables. Although in general the response rates for background variables were high, some variables and some countries exhibited less than acceptable rates. Since the student is the unit of analysis, the non-response rates given in the international report always reflect the percentage of students for whom the required responses from students, parents, teachers, or schools were not available. The following special notations were used to convey information about response rates in exhibits in the international report.

- For a country where student, parent, teacher or school responses were available for 70 percent to 84 percent of the students, an "r" appears next to the data for that country.

- When student, parent, teacher or school responses were available for 50 to 69 percent of the students, an "s" appears next to the data for that country.

- When student, parent, teacher or school responses were available for fewer than 50 percent of the students, an "x" replaces the data.

- When the percentage of students in a particular category fell below 2 percent, achievement data were not reported in that category. The data were replaced by a tilde (~).

- When data were unavailable for all respondents in a country, dashes (–) were used in place of data in all of the affected columns.

12.5.6 Development of the PIRLS International Report

The goal of the PIRLS international report was to describe fourth-grade students' reading achievement in participating countries and present as much information as possible about the contexts for learning to read. Beginning in September 2001, staff at the PIRLS International Study Center drafted an outline of the report, and, following a careful review of the questionnaires, developed specifications for the variables and indices to be included. Staff also prepared detailed analysis plans specifying how the analyses underlying each proposed exhibit in the draft report outline should be conducted, and began work developing the programs to implement the plans. Analysis plans included detailed documentation of the variables and response categories

involved, and the specification for any country-specific modifications to analyses necessitated by national adaptations to questions. These plans were incorporated in analysis notes for each proposed exhibit. The analyses required to produce the proposed exhibits were planned, and prototype exhibits prepared.

The analysis plans, report outlines, and prototype exhibits underwent a lengthy review involving the National Research Coordinators and project staff, following which consensus was achieved as to the contents of the international report, including the indices and variables to be reported. The analysis plans, outlines, and prototype exhibits were reviewed at the seventh meeting of the PIRLS 2001 National Research Coordinators in Athens, Greece, in March 2002. Following this meeting, the material was revised and updated to reflect the ideas and suggestions that were made. Some exhibits were deleted or added, and some of the analyses or presentational modes were modified.

After the data for all countries became available for analysis in mid-2002, the International Study Center conducted the psychometric scaling of the reading achievement data[7] and implemented the analyses documented in the analysis notes. In September 2002, staff met with the

PIRLS Reading Development Group to conduct scale-anchoring. Analyses were completed and the text of the report drafted in November 2002, after which draft reports were circulated by mail to NRCs for review. The draft report was reviewed in detail by NRCs at the eighth and final PIRLS NRC meeting in Istanbul, Turkey, in December 2002. Comments and suggestions from NRCs were incorporated into the final version of the report. Final revisions were made in January 2003, and the report was published in April 2003 (Mullis et al., 2003).

---

7   The scaling of the PIRLS achievement data is described in Chapter 11.

References

Gonzalez, E.G. & Gregory, K.D. (2000). Reporting student achievement in mathematics and science. In M.O. Martin, K.D. Gregory, & S.E. Stemler (Eds.), *TIMSS 1999 technical report.* Chestnut Hill, MA: Boston College.

Gregory, K.D. & Mullis, I.V.S. (2000). Describing international benchmarks of student achievement. In M.O. Martin, K.D. Gregory, & S.E. Stemler (Eds.), *TIMSS 1999 technical report.* Chestnut Hill, MA: Boston College.

Johnson, E.G., & Rust, K.F. (1992). Population references and variance estimation for NAEP data. *Journal of Educational Statistics,* 17, 175–190.

Kelly, D. L. (1999). *Interpreting the Third International Mathematics and Science Study (TIMSS) achievement scales using scale anchoring.* Unpublished doctoral dissertation, Boston College.

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., & Kennedy, A.M. (2003). *Trends in children's reading literacy achievement 1991–2002: IEA's repeat in nine countries of the 1991 Reading Literacy Study.* Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Kennedy, A.M. (2003). *PIRLS 2001 International Report: IEA's study of reading literacy achievement in primary schools in 35 countries.* Chestnut Hill, MA: Boston College.

Westat, Inc. (1997). *A user's guide to WesVarPC.* Rockville, MD: Westat, Inc.

Wolter, K.M. (1985). *Introduction to variance estimation.* New York: Springer-Verlag.