CHAPTER 4

Estimating Standard Errors in the PIRLS 2016 Results

Pierre Foy
Sylvie LaRoche

To obtain estimates of students' proficiency in reading that are both accurate and cost-effective, PIRLS 2016 made extensive use of probability sampling techniques to sample students from the national fourth grade student population, and applied matrix-sampling assessment designs to target individual students with a subset of the complete pool of assessment items. This approach made efficient use of resources, in particular keeping student response burden to a minimum, but at a cost of some variance or uncertainty in the reported statistics, such as the means and percentages computed to estimate population parameters.

To quantify this uncertainty, each statistic in the PIRLS 2016 and ePIRLS 2016 international reports is accompanied by an estimate of its standard error. For statistics reporting student achievement, which are based on plausible values, standard errors have two components. The first reflects the uncertainty due to generalizing from student samples to the entire fourth grade student population, referred to as sampling variance, and the second reflects uncertainty due to inferring students' performance on the entire assessment from their performance on the subset of items that they took, known as imputation variance. For parameter estimates of variables that are not plausible values, standard errors are based entirely on sampling variance.

Estimating Sampling Variance

PIRLS makes extensive use of probability sampling to derive achievement results from national samples of students. Because many such samples are possible but only one sample is drawn, some uncertainty about how well the sample represents the population is to be expected. The uncertainty caused by sampling students from a target population, known as sampling variance, can be estimated from the data of the one sample drawn.

Whereas estimating the sampling variance from simple random samples is a relatively easy task, estimating the sampling variance from the complex sample design of PIRLS is a more challenging endeavor.

A common way to estimate the sampling variance in multistage cluster sampling designs is through resampling schemes such as the balanced repeated replication and Jackknife techniques (Johnson & Rust, 1992; Wolter, 1985). PIRLS uses one variation of the Jackknife, the Jackknife Repeated Replication (JRR), to estimate sampling variances. JRR was chosen because it is computationally straightforward and provides approximately unbiased estimates of the sampling variances and sampling errors of means, total, and percentages.

At the core of the JRR technique is the grouping of primary sampling units—schools—into zones based on the sample design stratification and subsequent repeated draws of subsamples from these zones, i.e., repeated replication. For PIRLS, the two main features of the PIRLS sample design that JRR incorporates in its repeated draws of subsamples are the stratification of schools and the clustering of students within schools. This is done by defining Jackknife sampling zones according to the stratification scheme, pairing successive schools[1] to model the clustering from each national sample (see Chapter 3 for information on the Sample Design). Since most national samples consist of 150 schools, a total of 75 zones are created. If more than 150 schools are selected, then the additional zones are collapsed into the first 75 zones. The subsampling required by JRR is applied within each sampling zone.

Sampling zones are constructed within explicit strata. When an explicit stratum has an odd number of schools, either by design or because of school non-response, the students in the remaining school are randomly divided to make up two "quasi" schools.[2] Each sampling zone then consists of a pair of schools or "quasi" schools.

Exhibit 4.1 lists the number of sampling zones for each PIRLS and ePIRLS 2016 participating country.

1 When schools are sampled, schools are ordered within explicit strata by implicit stratification variables and the measure of size. Based on this sorting, successively sampled schools are matched and classified together in each sampling zone. More information can be found in Appendix 3A of Chapter 3.

2 For example, if a remaining school consists of 2 sampled classrooms, each classroom becomes a "quasi" school.

Exhibit 4.1: Number of Sampling Zones for Each PIRLS 2016 Participating Country

Country	PIRLS and ePIRLS 2016 Sampling Zones	
	PIRLS	ePIRLS
Australia	75	-
Austria	75	-
Azerbaijan	75	-
Bahrain	75	-
Belgium (Flemish)	75	-
Belgium (French)	75	-
Bulgaria	75	-
Canada	75	75
Chile	75	-
Chinese Taipei	75	75
Czech Republic	75	-
Denmark	75	72
Egypt	75	-
England	75	-
Finland	75	-
France	75	-
Georgia	75	75
Germany	75	-
Hong Kong SAR	70	-
Hungary	75	-
Iran, Islamic Rep. of	75	-
Ireland	75	74
Israel	75	75
Italy	75	75
Kazakhstan	75	-
Kuwait	75	-
Latvia	74	-
Lithuania	75	-
Macao SAR	75	-
Malta	75	-
Morocco	75	-
Netherlands	68	-
New Zealand	75	-
Northern Ireland	70	-
Norway (5)	75	71
Oman	75	-

Exhibit 4.1: Number of Sampling Zones for Each PIRLS 2016 Participating Country (Continued)

Country	PIRLS and ePIRLS 2016 Sampling Zones	
	PIRLS	ePIRLS
Poland	75	-
Portugal	75	75
Qatar	75	-
Russian Federation	75	-
Saudi Arabia	75	-
Singapore	75	75
Slovak Republic	75	-
Slovenia	75	75
South Africa	75	-
Spain	75	-
Sweden	75	73
Trinidad and Tobago	75	-
United Arab Emirates	75	75
United States	75	75
Benchmarking Participants		
Buenos Aires, Argentina	75	-
Ontario, Canada	75	-
Quebec, Canada	65	-
Denmark (3)	75	-
Norway (4)	75	-
Moscow City, Russian Fed.	75	-
Eng/Afr/Zulu - RSA (5)	64	-
Andalusia, Spain	75	-
Madrid, Spain	75	-
Abu Dhabi, UAE	72	73
Dubai, UAE	75	75

The JRR procedure draws two subsamples from each sampling zone: one where the first school in the pair is included and the second school is removed, and another subsample where the second school is included and the first school is removed. When a school is removed from the sample, the weights of the remaining school are doubled to make up for the omitted school. In both subsamples, all students in the other sampling zones are included. With this process applied in each of the 75 sampling zones, the JRR procedure yields a total of 150 replicate subsamples, each

one with its own set of replicate sampling weights to account for the successive removal of each school from the pair of schools in any given sampling zone.[3]

The process of creating replicate sampling weights for the replicate subsamples defines replicate factors k_{hj} as follows:

$$k_{hj} = \begin{cases} 2 \text{ for students in school } j \text{ of sampling zone } h \\ 0 \text{ for students in the other school of sampling zone } h \\ 1 \text{ for students in any other sampling zone} \end{cases} \quad (4.1)$$

These replicate factors are used to compute the 150 sets of replicate sampling weights as follows:

$$W_{hji} = k_{hj} \cdot W_{0i} \quad (4.2)$$

where W_{0i} is the overall sampling weight of student i and W_{hji} is the resulting replicate sampling weight of student i from sampling zone h when school j is included and the other school in the pair is removed.

Exhibit 4.2 illustrates how the replicate factors, necessary to produce the replicate sampling weights, are derived. Within each sampling zone, each school is assigned randomly an indicator u_{hj}, coded either 0 or 1, such that one school has a value of 0 and the other a value of 1. This indicator serves to identify which schools within each zone will be successively included or removed. When a school is removed from a zone, the replicate factor is set to zero and the sampling weights of all students in that school are set to zero; when a school is included, the replicate factor is set to two and the sampling weights of all students in that school are doubled. The sampling weights of students in all other sampling zones remain unchanged.

For example, sampling zone 1 yields two sets of replicate sampling weights. The first set has doubled sampling weights ($k_{11} = 2$) for the students in the first school ($u_{11} = 0$) of zone 1, zeroed sampling weights ($k_{12} = 0$) for the students in the second school ($u_{12} = 1$) of zone 1, and unchanged sampling weights ($k_{hj} = 1$) for all students in the other sampling zones. The second set of replicate sampling weights has zeroed sampling weights ($k_{11} = 0$) for the students in the first school ($u_{11} = 0$) of zone 1, doubled sampling weights ($k_{12} = 2$) for the students in the second school ($u_{12} = 1$) of zone 1, and unchanged sampling weights ($k_{hj} = 1$) for all students in the other sampling zones.

3 Prior to 2016, PIRLS used 75 subsamples and sets of replicate weights to calculate the JRR sampling variances. To provide more accurate estimates, starting in 2016 PIRLS uses 150 subsamples and sets of replicate weights to calculate the JRR sampling variances. Two subsamples are drawn from each sampling zone rather than one randomly selected subsample.

Exhibit 4.2: Construction of Replicate Factors Across Sampling Zones

Sample Zone	School Replicate Indicator (u_{hj})	Replicate Factors for Computing JRR Replicate Sampling Weights (k_{hj})											
		Zone 1		Zone 2		Zone 3		…	Zone h		…	Zone 75	
		(1)	(2)	(3)	(4)	(5)	(6)		(2h–1)	(2h)		(149)	(150)
1	0	2	0	1	1	1	1	…	1	1	…	1	1
	1	0	2										
2	0	1	1	2	0	1	1	…	1	1	…	1	1
	1			0	2								
3	0	1	1	1	1	2	0	…	1	1	…	1	1
	1					0	2						
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋮	⋮
h	0	1	1	1	1	1	1	…	2	0	…	1	1
	1								0	2			
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮
75	0	1	1	1	1	1	1	…	1	1	…	2	0
	1											0	2

The process is repeated across all 75 possible sampling zones, generating 150 sets of replicate sampling weights. The replicate sampling weights are then used to estimate a statistic of interest 150 times. The variation across these 150 jackknife estimates determines the sampling variance.

Given a statistic t to be computed from a national sample, the formula used to estimate the sampling variance of that statistic, based on the PIRLS JRR algorithm, is given by the following equation:

$$Var_{jrr}(t_0) = \frac{1}{2}\sum_{h=1}^{75}\sum_{j=1}^{2}(t_{hj} - t_0)^2 \qquad (4.3)$$

where the term t_0 denotes the statistic of interest estimated with the overall student sampling weights W_{0i} and the term t_{hj} denotes the same statistic computed using the set of replicate sampling weights W_{hji} obtained from sampling zone h (h=1,...,75), where the j^{th} school (1st or 2nd) in the zone is included and the other removed.

The sampling variance estimated with the PIRLS JRR method represents the variation arising from having sampled students using the multi-stage stratified cluster sample design. Its square root is the standard error for any statistic derived from variables other than plausible values. Examples of such statistics include the mean age of students, the mean scale score on the PIRLS *Students Like Reading* contextual scale, and the percentage of students that attended preprimary education three years or more.

Estimating Imputation Variance

For variables other than plausible values, standard errors were the result solely of sampling variation, and were computed using the JRR technique. However, the situation for plausible values was more complicated. As described in Chapter 2 of the *PIRLS 2016 Assessment Framework*, the PIRLS item pool was far too extensive to be administered in its entirety to any one student, and so a matrix-sampling assessment design was adopted whereby each student was given a single test booklet containing only a part of the entire assessment. The results for all of the booklets were then aggregated using item response theory (IRT) to provide results for the entire assessment. Multiple imputation was used to derive reliable estimates of student performance (plausible values) on the assessment as a whole, even though each student responded to just a subset of the assessment items. Because every student proficiency estimate incorporates a random element, PIRLS 2016 followed the customary procedure of generating five estimates for each student and using the variability among them as a measure of the imputation uncertainty, or error.

The general procedure for estimating the imputation variance when analyzing student achievement data follows the basic principle of performing any statistical analysis five times—once for each set of plausible values—and aggregating the five sets of results (Mislevy et al., 1992). Thus, for any given achievement-based statistic t, estimating that statistic from each plausible value yields five estimates t_m, $m = 1, \ldots, 5$, all of them computed using the overall student sampling weights W_{0i}. The final estimate of that statistic, t_0, is the average of these five estimates:

$$t_0 = \frac{1}{5}\sum_{m=1}^{5} t_m \tag{4.4}$$

The imputation variance of the statistic t_0 is simply the variance of the five results from the plausible values, computed as follows:

$$Var_{imp}(t_0) = \frac{6}{5}\sum_{m=1}^{5} \frac{(t_m - t_0)^2}{4} \tag{4.5}$$

where the factor $\frac{6}{5}$ is a correction factor required by the multiple imputation methodology. This imputation variance is then added to the sampling variance to produce the total variance estimate of the statistic t_0, as follows:

$$Var_{tot}(t_0) = Var_{jrr}(t_0) + Var_{imp}(t_0) \tag{4.6}$$

The sampling variance in this context is the average of the sampling variances from the five plausible values, as follows:

$$Var_{jrr}(t_0) = \frac{1}{5}\sum_{m=1}^{5} Var_{jrr}(t_m) \tag{4.7}$$

where

$$Var_{jrr}(t_m) = \frac{1}{2}\sum_{h=1}^{75}\sum_{j=1}^{2}(t_{mhj} - t_m)^2 \tag{4.8}$$

and t_{mhj} is the appropriate JRR estimate based on plausible value computed using the set of replicate sampling weights from sampling zone h where school j is included. The square root of the total variance is then the proper standard error for any statistic based on plausible values, such as the average PIRLS reading achievement for girls or the percentage of students who reached the PIRLS Advanced International Benchmark of reading achievement.

Appendices 4A and 4B provide details on the jackknife sampling variance, the imputation variance, the total variance, and the overall standard error for each country's mean proficiency estimates for PIRLS 2016 and ePIRLS 2016, respectively.

Estimating Standard Errors for International Averages

Some exhibits in the PIRLS 2016 reports include international averages and their standard errors, listed at the bottom of the exhibit. For example, Exhibit 1.5 of the *PIRLS 2016 International Results in Reading* report provides the international average for the percentages of girls and boys and their fourth grade reading achievement at the bottom of the exhibit. International averages are computed using the data from the participating countries included in the main table of an exhibit. Data from the benchmarking participants are not included in the estimation of international averages.

For any given statistic t_0, its international average is given by:

$$t_{int} = \frac{1}{N}\sum_{i=1}^{N} t_{0i} \tag{4.9}$$

where N is the number of countries contributing to the international average and t_{0i} is the estimate of our statistic of interest for the i^{th} country.

The variance of the international t_{int} average is given by:

$$Var(t_{int}) = \frac{1}{N^2}\sum_{i=1}^{N} Var_{tot}(t_{0i}) \tag{4.10}$$

where $Var_{tot}(t_{0i})$ is the total variance of our statistic of interest for the i^{th} country, as given in Equation (4.6) above. For statistics based on plausible values, the total variance includes the sampling variance and the imputation variance. For statistics not based on plausible values, such as percentages, the total variance is based entirely on the sampling variance, as shown in Equation (4.3) above. The standard error of the international average is the square root of the total variance.

References

Johnson, E.G., & Rust, K.F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics, 17*(2), 175–190.

Mislevy, R.J., Beaton, A., Kaplan, B.A., & Sheehan, K. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement, 29*(2), 133–161.

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

Appendix 4A: Summary Statistics and Standard Errors for Proficiency in PIRLS Reading

Summary Statistics and Standard Errors for Proficiency in Overall Reading

Country	Sample Size	Overall Reading				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	6,341	544.360	5.802	0.614	6.416	2.533
Austria	4,360	540.796	5.297	0.370	5.667	2.381
Azerbaijan	5,994	472.277	16.798	0.650	17.447	4.177
Bahrain	5,480	445.999	4.549	0.854	5.403	2.325
Belgium (Flemish)	5,198	525.059	3.415	0.338	3.752	1.937
Belgium (French)	4,623	497.495	4.912	1.696	6.608	2.571
Bulgaria	4,281	551.539	17.215	0.794	18.009	4.244
Canada	18,245	543.098	3.182	0.165	3.348	1.830
Chile	4,294	493.872	5.408	0.812	6.220	2.494
Chinese Taipei	4,326	558.894	3.442	0.714	4.155	2.038
Czech Republic	5,537	543.348	4.245	0.244	4.488	2.119
Denmark	3,508	547.492	3.974	0.494	4.468	2.114
Egypt	6,957	330.471	28.680	3.194	31.874	5.646
England	5,095	558.682	3.461	0.117	3.578	1.891
Finland	4,896	566.007	3.090	0.311	3.400	1.844
France	4,767	511.244	4.510	0.225	4.736	2.176
Georgia	5,741	488.319	7.131	0.752	7.883	2.808
Germany	3,959	537.325	9.657	0.395	10.052	3.170
Hong Kong SAR	3,349	568.583	7.068	0.470	7.538	2.746
Hungary	4,623	554.160	8.178	0.114	8.292	2.880
Iran, Islamic Rep. of	8,766	427.899	13.543	2.065	15.608	3.951
Ireland	4,607	566.596	6.091	0.049	6.139	2.478
Israel	4,041	530.288	6.044	0.410	6.454	2.541
Italy	3,940	548.007	3.691	1.073	4.765	2.183
Kazakhstan	4,925	536.046	5.792	0.374	6.166	2.483
Kuwait	4,609	393.432	15.627	1.510	17.137	4.140
Latvia	4,157	557.751	2.724	0.158	2.882	1.698
Lithuania	4,317	548.278	5.969	0.929	6.898	2.626
Macao SAR	4,059	545.581	0.741	0.337	1.077	1.038
Malta	3,647	452.012	1.711	1.557	3.269	1.808

Summary Statistics and Standard Errors for Proficiency in Overall Reading (Continued)

Country	Sample Size	Overall Reading				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Morocco	10,942	357.823	13.813	1.694	15.506	3.938
Netherlands	4,206	544.884	2.670	0.132	2.803	1.674
New Zealand	5,646	522.531	4.517	0.224	4.741	2.177
Northern Ireland	3,693	564.621	4.201	0.701	4.902	2.214
Norway (5)	4,232	558.950	4.767	0.333	5.100	2.258
Oman	9,234	418.483	10.301	0.772	11.073	3.328
Poland	4,413	564.626	4.256	0.234	4.490	2.119
Portugal	4,642	527.797	4.944	0.254	5.198	2.280
Qatar	9,077	442.246	2.758	0.656	3.414	1.848
Russian Federation	4,577	580.772	4.397	0.448	4.846	2.201
Saudi Arabia	4,741	430.300	16.339	1.234	17.573	4.192
Singapore	6,488	576.178	9.769	0.171	9.940	3.153
Slovak Republic	5,451	534.791	8.555	1.163	9.719	3.117
Slovenia	4,499	542.466	3.640	0.267	3.907	1.977
South Africa	12,810	319.629	18.499	1.027	19.525	4.419
Spain	14,595	527.740	2.871	0.124	2.995	1.731
Sweden	4,525	555.160	5.550	0.195	5.745	2.397
Trinidad and Tobago	4,177	479.404	9.996	0.723	10.720	3.274
United Arab Emirates	16,471	450.104	9.815	0.641	10.456	3.234
United States	4,425	549.441	8.741	0.806	9.548	3.090
Benchmarking Participants						
Buenos Aires, Argentina	4,382	479.957	8.949	0.440	9.390	3.064
Ontario, Canada	4,270	543.582	10.050	0.225	10.276	3.206
Quebec, Canada	3,179	547.422	7.881	0.198	8.079	2.842
Denmark (3)	3,600	500.875	4.543	2.701	7.243	2.691
Norway (4)	4,354	516.874	3.283	0.593	3.876	1.969
Moscow City, Russian Fed.	4,289	612.084	4.170	0.476	4.646	2.155
Eng/Afr/Zulu - RSA (5)	5,282	406.012	34.276	1.391	35.667	5.972
Andalusia, Spain	4,169	524.584	4.113	0.232	4.345	2.084
Madrid, Spain	3,794	549.014	3.305	0.660	3.966	1.991
Abu Dhabi, UAE	4,188	414.308	20.632	1.567	22.199	4.712
Dubai, UAE	7,859	514.992	3.330	0.279	3.609	1.900

Summary Statistics and Standard Errors for Proficiency in Literary Experience

Country	Sample Size	Literary Experience				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	6,341	547.205	5.474	0.413	5.887	2.426
Austria	4,360	544.303	5.140	0.285	5.425	2.329
Azerbaijan	5,994	465.851	14.752	0.631	15.383	3.922
Bahrain	5,480	437.478	6.059	1.920	7.980	2.825
Belgium (Flemish)	5,198	523.594	3.637	0.111	3.748	1.936
Belgium (French)	4,623	503.821	4.786	0.139	4.925	2.219
Bulgaria	4,281	551.441	19.008	0.937	19.945	4.466
Canada	18,245	547.215	3.309	0.153	3.462	1.861
Chile	4,294	500.389	5.670	0.611	6.281	2.506
Chinese Taipei	4,326	548.387	3.265	0.907	4.172	2.042
Czech Republic	5,537	544.982	4.171	0.282	4.452	2.110
Denmark	3,508	551.281	3.972	0.709	4.681	2.163
Egypt	6,957	328.138	29.034	1.327	30.360	5.510
England	5,095	562.603	3.697	1.104	4.801	2.191
Finland	4,896	564.900	3.292	0.276	3.568	1.889
France	4,767	512.680	5.034	0.765	5.799	2.408
Georgia	5,741	489.905	6.860	0.071	6.931	2.633
Germany	3,959	542.338	9.951	0.875	10.826	3.290
Hong Kong SAR	3,349	562.473	7.404	1.505	8.909	2.985
Hungary	4,623	557.611	7.959	0.142	8.102	2.846
Iran, Islamic Rep. of	8,766	430.257	13.464	1.185	14.649	3.827
Ireland	4,607	571.308	6.148	1.111	7.259	2.694
Israel	4,041	532.226	6.399	0.291	6.690	2.587
Italy	3,940	548.737	4.037	0.406	4.442	2.108
Kazakhstan	4,925	527.236	5.512	0.810	6.322	2.514
Kuwait	4,609	387.778	16.543	1.820	18.363	4.285
Latvia	4,157	555.030	2.802	0.824	3.626	1.904
Lithuania	4,317	547.418	5.620	1.797	7.417	2.723
Macao SAR	4,059	535.999	0.892	1.954	2.846	1.687
Malta	3,647	451.899	2.075	1.890	3.965	1.991
Morocco	10,942	353.248	14.589	1.775	16.364	4.045
Netherlands	4,206	546.355	2.793	0.066	2.858	1.691

Summary Statistics and Standard Errors for Proficiency in Literary Experience (Continued)

Country	Sample Size	Literary Experience				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
New Zealand	5,646	525.278	4.731	0.650	5.381	2.320
Northern Ireland	3,693	570.464	4.563	1.599	6.162	2.482
Norway (5)	4,232	560.337	4.833	1.589	6.422	2.534
Oman	9,234	410.713	10.122	1.058	11.180	3.344
Poland	4,413	566.610	4.066	0.823	4.889	2.211
Portugal	4,642	527.774	5.208	0.906	6.115	2.473
Qatar	9,077	434.112	2.927	2.139	5.067	2.251
Russian Federation	4,577	579.129	4.237	0.471	4.708	2.170
Saudi Arabia	4,741	429.967	14.684	1.169	15.852	3.982
Singapore	6,488	574.559	9.542	1.193	10.735	3.276
Slovak Republic	5,451	538.758	8.747	0.407	9.154	3.026
Slovenia	4,499	541.192	3.653	2.152	5.805	2.409
South Africa	12,810	323.042	18.590	3.526	22.116	4.703
Spain	14,595	530.000	3.301	0.184	3.485	1.867
Sweden	4,525	555.953	5.513	0.231	5.744	2.397
Trinidad and Tobago	4,177	478.242	10.113	0.949	11.062	3.326
United Arab Emirates	16,471	439.953	10.606	0.638	11.244	3.353
United States	4,425	557.260	8.998	0.155	9.153	3.025
Benchmarking Participants						
Buenos Aires, Argentina	4,382	483.764	8.924	0.837	9.761	3.124
Ontario, Canada	4,270	548.600	10.074	0.398	10.473	3.236
Quebec, Canada	3,179	549.563	7.784	0.704	8.488	2.913
Denmark (3)	3,600	504.870	4.408	1.877	6.285	2.507
Norway (4)	4,354	520.498	3.437	0.783	4.220	2.054
Moscow City, Russian Fed.	4,289	613.262	4.095	0.915	5.010	2.238
Eng/Afr/Zulu - RSA (5)	5,282	401.912	37.153	2.567	39.720	6.302
Andalusia, Spain	4,169	525.589	4.068	0.274	4.342	2.084
Madrid, Spain	3,794	550.505	3.790	1.004	4.795	2.190
Abu Dhabi, UAE	4,188	405.509	22.119	0.938	23.058	4.802
Dubai, UAE	7,859	507.966	3.473	0.841	4.314	2.077

Summary Statistics and Standard Errors for Proficiency in Acquire and Use Information

Country	Sample Size	Acquire and Use Information				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	6,341	542.524	6.557	0.282	6.838	2.615
Austria	4,360	538.867	5.260	0.463	5.723	2.392
Azerbaijan	5,994	477.356	20.516	1.031	21.547	4.642
Bahrain	5,480	453.124	3.610	0.999	4.609	2.147
Belgium (Flemish)	5,198	525.831	3.297	0.431	3.728	1.931
Belgium (French)	4,623	490.108	5.189	0.429	5.619	2.370
Bulgaria	4,281	553.851	16.750	0.714	17.464	4.179
Canada	18,245	540.080	3.259	0.422	3.681	1.919
Chile	4,294	485.055	6.274	0.755	7.029	2.651
Chinese Taipei	4,326	569.214	3.530	1.157	4.687	2.165
Czech Republic	5,537	541.247	4.388	0.773	5.162	2.272
Denmark	3,508	543.284	4.565	1.737	6.302	2.510
Egypt	6,957	331.918	29.623	3.833	33.455	5.784
England	5,095	556.423	3.946	0.306	4.252	2.062
Finland	4,896	568.741	3.637	0.191	3.828	1.956
France	4,767	510.087	4.748	0.868	5.616	2.370
Georgia	5,741	486.383	8.775	0.990	9.765	3.125
Germany	3,959	532.921	9.859	1.197	11.056	3.325
Hong Kong SAR	3,349	576.386	7.236	0.358	7.594	2.756
Hungary	4,623	550.557	9.239	1.566	10.805	3.287
Iran, Islamic Rep. of	8,766	424.585	14.312	0.424	14.737	3.839
Ireland	4,607	564.727	6.584	0.454	7.038	2.653
Israel	4,041	528.681	6.168	0.044	6.212	2.492
Italy	3,940	548.960	3.737	1.203	4.940	2.223
Kazakhstan	4,925	543.594	6.649	1.361	8.010	2.830
Kuwait	4,609	398.428	18.028	0.489	18.517	4.303
Latvia	4,157	561.315	2.881	0.345	3.227	1.796
Lithuania	4,317	550.574	5.983	0.884	6.867	2.620
Macao SAR	4,059	555.505	0.779	0.892	1.671	1.293
Malta	3,647	451.399	1.438	2.558	3.996	1.999
Morocco	10,942	358.695	14.557	1.311	15.868	3.983
Netherlands	4,206	544.693	3.118	0.479	3.597	1.897

Summary Statistics and Standard Errors for Proficiency in Acquire and Use Information (Continued)

Country	Sample Size	Acquire and Use Information				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
New Zealand	5,646	520.376	4.869	0.851	5.720	2.392
Northern Ireland	3,693	560.782	4.273	1.094	5.368	2.317
Norway (5)	4,232	558.569	5.180	0.465	5.644	2.376
Oman	9,234	425.473	10.333	0.781	11.114	3.334
Poland	4,413	564.334	5.147	1.615	6.762	2.600
Portugal	4,642	528.388	5.065	0.335	5.400	2.324
Qatar	9,077	449.632	2.759	0.858	3.617	1.902
Russian Federation	4,577	584.419	4.669	0.398	5.068	2.251
Saudi Arabia	4,741	428.825	19.915	0.630	20.545	4.533
Singapore	6,488	578.591	10.208	0.724	10.932	3.306
Slovak Republic	5,451	531.052	9.387	0.399	9.786	3.128
Slovenia	4,499	544.294	3.864	0.459	4.323	2.079
South Africa	12,810	313.765	18.933	0.947	19.880	4.459
Spain	14,595	526.599	2.347	0.211	2.558	1.599
Sweden	4,525	554.850	6.242	0.550	6.792	2.606
Trinidad and Tobago	4,177	479.947	9.994	2.259	12.252	3.500
United Arab Emirates	16,471	459.769	9.255	0.958	10.214	3.196
United States	4,425	543.084	9.330	0.258	9.588	3.096
Benchmarking Participants						
Buenos Aires, Argentina	4,382	475.330	9.153	1.589	10.742	3.278
Ontario, Canada	4,270	539.458	11.135	0.659	11.794	3.434
Quebec, Canada	3,179	546.662	8.664	0.446	9.111	3.018
Denmark (3)	3,600	497.789	5.268	0.414	5.683	2.384
Norway (4)	4,354	513.681	3.523	1.179	4.701	2.168
Moscow City, Russian Fed.	4,289	613.081	4.901	1.550	6.450	2.540
Eng/Afr/Zulu - RSA (5)	5,282	407.024	35.042	0.725	35.766	5.981
Andalusia, Spain	4,169	523.891	4.030	0.831	4.860	2.205
Madrid, Spain	3,794	548.969	3.382	0.505	3.887	1.971
Abu Dhabi, UAE	4,188	422.034	19.937	4.633	24.570	4.957
Dubai, UAE	7,859	523.258	3.060	1.341	4.401	2.098

Summary Statistics and Standard Errors for Proficiency in Retrieving and Straightforward Inferencing

Country	Sample Size	Retrieving and Straightforward Inferencing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	6,341	540.737	5.620	0.979	6.599	2.569
Austria	4,360	550.244	6.319	1.601	7.920	2.814
Azerbaijan	5,994	477.348	16.328	1.155	17.483	4.181
Bahrain	5,480	444.491	4.017	0.462	4.478	2.116
Belgium (Flemish)	5,198	525.738	3.546	0.958	4.503	2.122
Belgium (French)	4,623	500.618	4.830	0.361	5.192	2.279
Bulgaria	4,281	550.435	16.093	0.103	16.196	4.024
Canada	18,245	541.431	3.109	0.106	3.215	1.793
Chile	4,294	495.873	5.478	0.606	6.084	2.466
Chinese Taipei	4,326	559.843	3.216	0.504	3.720	1.929
Czech Republic	5,537	551.194	4.669	1.051	5.720	2.392
Denmark	3,508	549.832	4.482	0.055	4.537	2.130
Egypt	6,957	329.082	29.228	2.052	31.280	5.593
England	5,095	555.703	3.678	0.305	3.983	1.996
Finland	4,896	572.066	3.660	0.249	3.908	1.977
France	4,767	520.580	4.772	0.477	5.249	2.291
Georgia	5,741	486.107	6.777	0.236	7.013	2.648
Germany	3,959	545.641	10.617	0.146	10.763	3.281
Hong Kong SAR	3,349	567.558	6.268	1.128	7.396	2.720
Hungary	4,623	551.523	8.892	1.821	10.714	3.273
Iran, Islamic Rep. of	8,766	429.490	14.304	1.796	16.100	4.012
Ireland	4,607	566.023	6.413	0.504	6.917	2.630
Israel	4,041	529.722	5.490	0.500	5.991	2.448
Italy	3,940	546.713	3.861	0.440	4.300	2.074
Kazakhstan	4,925	529.348	5.982	0.295	6.277	2.505
Kuwait	4,609	393.889	15.589	1.190	16.779	4.096
Latvia	4,157	554.060	2.938	0.786	3.724	1.930
Lithuania	4,317	549.379	5.557	0.945	6.503	2.550
Macao SAR	4,059	549.143	0.770	0.451	1.221	1.105
Malta	3,647	451.850	1.708	1.283	2.991	1.730
Morocco	10,942	363.775	13.018	2.071	15.089	3.884
Netherlands	4,206	546.451	3.070	0.929	3.999	2.000

Summary Statistics and Standard Errors for Proficiency in Retrieving and Straightforward Inferencing (Continued)

Country	Sample Size	Retrieving and Straightforward Inferencing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
New Zealand	5,646	521.368	4.408	1.014	5.421	2.328
Northern Ireland	3,693	561.558	4.251	0.304	4.554	2.134
Norway (5)	4,232	561.465	4.462	1.065	5.527	2.351
Oman	9,234	419.474	9.287	1.213	10.500	3.240
Poland	4,413	559.706	4.223	0.297	4.520	2.126
Portugal	4,642	527.791	4.700	0.243	4.943	2.223
Qatar	9,077	442.096	2.561	0.688	3.250	1.803
Russian Federation	4,577	581.389	4.941	0.210	5.152	2.270
Saudi Arabia	4,741	425.246	15.605	0.972	16.577	4.071
Singapore	6,488	573.013	9.375	0.274	9.649	3.106
Slovak Republic	5,451	537.543	8.881	0.533	9.414	3.068
Slovenia	4,499	546.631	4.628	0.435	5.063	2.250
South Africa	12,810	321.276	17.988	2.184	20.172	4.491
Spain	14,595	526.460	2.683	0.200	2.883	1.698
Sweden	4,525	560.141	6.157	0.997	7.154	2.675
Trinidad and Tobago	4,177	483.498	9.629	3.062	12.691	3.562
United Arab Emirates	16,471	448.078	9.133	1.151	10.283	3.207
United States	4,425	542.892	8.225	0.840	9.065	3.011
Benchmarking Participants						
Buenos Aires, Argentina	4,382	482.869	7.501	0.854	8.355	2.890
Ontario, Canada	4,270	538.853	9.940	0.749	10.689	3.269
Quebec, Canada	3,179	550.986	8.265	0.974	9.239	3.040
Denmark (3)	3,600	500.102	4.890	0.514	5.404	2.325
Norway (4)	4,354	521.395	3.411	0.617	4.028	2.007
Moscow City, Russian Fed.	4,289	611.229	4.378	1.182	5.559	2.358
Eng/Afr/Zulu - RSA (5)	5,282	407.415	34.318	3.110	37.428	6.118
Andalusia, Spain	4,169	522.016	3.549	0.081	3.630	1.905
Madrid, Spain	3,794	546.754	3.751	0.181	3.932	1.983
Abu Dhabi, UAE	4,188	412.982	19.278	1.499	20.776	4.558
Dubai, UAE	7,859	511.647	2.886	2.777	5.663	2.380

Summary Statistics and Standard Errors for Proficiency in Interpreting, Integrating, and Evaluating

Country	Sample Size	Interpreting, Integrating, and Evaluating				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	6,341	549.233	5.487	0.116	5.602	2.367
Austria	4,360	534.439	5.227	1.174	6.401	2.530
Azerbaijan	5,994	464.716	18.646	0.251	18.896	4.347
Bahrain	5,480	445.507	4.668	2.651	7.319	2.705
Belgium (Flemish)	5,198	524.358	3.600	1.283	4.883	2.210
Belgium (French)	4,623	494.123	5.098	0.718	5.816	2.412
Bulgaria	4,281	552.315	18.098	0.383	18.481	4.299
Canada	18,245	545.111	3.146	0.204	3.350	1.830
Chile	4,294	491.144	6.137	2.357	8.494	2.914
Chinese Taipei	4,326	558.093	3.729	1.074	4.803	2.192
Czech Republic	5,537	537.904	4.207	0.648	4.855	2.203
Denmark	3,508	546.073	3.348	1.311	4.660	2.159
Egypt	6,957	339.914	26.558	6.019	32.577	5.708
England	5,095	561.489	3.453	0.156	3.609	1.900
Finland	4,896	562.473	3.246	0.054	3.300	1.817
France	4,767	501.030	5.169	0.376	5.545	2.355
Georgia	5,741	489.857	7.743	0.538	8.280	2.878
Germany	3,959	530.124	10.174	0.027	10.201	3.194
Hong Kong SAR	3,349	568.476	7.786	0.579	8.365	2.892
Hungary	4,623	556.735	8.693	0.367	9.060	3.010
Iran, Islamic Rep. of	8,766	424.686	13.694	2.953	16.647	4.080
Ireland	4,607	569.284	6.732	1.634	8.366	2.892
Israel	4,041	530.049	6.750	0.764	7.514	2.741
Italy	3,940	549.634	3.691	0.707	4.398	2.097
Kazakhstan	4,925	542.378	5.531	0.183	5.714	2.390
Kuwait	4,609	388.427	17.284	2.985	20.269	4.502
Latvia	4,157	561.800	2.574	0.239	2.813	1.677
Lithuania	4,317	547.824	5.935	1.025	6.960	2.638
Macao SAR	4,059	543.009	0.753	1.832	2.585	1.608
Malta	3,647	451.124	1.957	1.529	3.486	1.867
Morocco	10,942	336.140	17.391	2.660	20.052	4.478
Netherlands	4,206	544.387	2.649	0.359	3.008	1.734

Summary Statistics and Standard Errors for Proficiency in Interpreting, Integrating, and Evaluating (Continued)

Country	Sample Size	Interpreting, Integrating, and Evaluating				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
New Zealand	5,646	524.565	4.821	0.793	5.613	2.369
Northern Ireland	3,693	567.411	4.291	0.598	4.889	2.211
Norway (5)	4,232	558.138	4.888	0.879	5.767	2.401
Oman	9,234	414.785	11.307	1.473	12.781	3.575
Poland	4,413	569.549	4.132	1.472	5.604	2.367
Portugal	4,642	526.449	5.421	0.279	5.701	2.388
Qatar	9,077	440.920	2.627	0.918	3.544	1.883
Russian Federation	4,577	582.051	4.043	0.898	4.941	2.223
Saudi Arabia	4,741	438.563	16.027	1.174	17.200	4.147
Singapore	6,488	578.805	9.880	0.096	9.976	3.159
Slovak Republic	5,451	531.427	8.659	1.572	10.231	3.199
Slovenia	4,499	539.426	4.162	1.979	6.141	2.478
South Africa	12,810	308.245	20.834	6.739	27.572	5.251
Spain	14,595	529.100	2.729	0.251	2.979	1.726
Sweden	4,525	552.831	5.875	0.428	6.303	2.511
Trinidad and Tobago	4,177	472.284	11.842	1.377	13.219	3.636
United Arab Emirates	16,471	452.586	9.924	1.018	10.942	3.308
United States	4,425	554.964	8.973	0.651	9.624	3.102
Benchmarking Participants						
Buenos Aires, Argentina	4,382	472.816	11.927	1.640	13.567	3.683
Ontario, Canada	4,270	548.168	9.982	0.050	10.032	3.167
Quebec, Canada	3,179	545.009	8.457	0.526	8.983	2.997
Denmark (3)	3,600	503.905	5.126	1.026	6.152	2.480
Norway (4)	4,354	512.696	3.568	0.204	3.772	1.942
Moscow City, Russian Fed.	4,289	614.113	4.081	0.198	4.278	2.068
Eng/Afr/Zulu - RSA (5)	5,282	399.808	33.851	5.129	38.980	6.243
Andalusia, Spain	4,169	526.894	4.186	1.151	5.337	2.310
Madrid, Spain	3,794	549.928	3.759	0.129	3.888	1.972
Abu Dhabi, UAE	4,188	416.586	20.209	2.267	22.475	4.741
Dubai, UAE	7,859	518.784	3.165	0.469	3.634	1.906

Appendix 4B: Summary Statistics and Standard Errors for Proficiency in ePIRLS Online Informational Reading

Summary Statistics and Standard Errors for Proficiency in Overall ePIRLS Online Informational Reading

Country	Sample Size	Overall ePIRLS Online Informational Reading				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Canada	8,871	542.622	9.597	0.436	10.033	3.168
Chinese Taipei	4,299	545.648	3.651	0.279	3.930	1.982
Denmark	2,506	558.288	4.552	0.404	4.956	2.226
Georgia	5,557	476.903	10.138	0.613	10.752	3.279
Ireland	2,473	566.799	6.315	0.175	6.489	2.547
Israel	3,798	536.134	5.009	0.480	5.489	2.343
Italy	3,767	532.465	4.227	0.370	4.597	2.144
Norway (5)	3,610	567.537	4.316	0.597	4.913	2.217
Portugal	4,558	522.386	4.174	0.874	5.048	2.247
Singapore	6,320	588.129	9.024	0.147	9.171	3.028
Slovenia	4,303	525.010	3.153	0.649	3.802	1.950
Sweden	3,879	559.204	5.265	0.108	5.373	2.318
United Arab Emirates	15,566	468.330	4.365	0.475	4.840	2.200
United States	4,090	556.552	6.582	0.142	6.724	2.593
Benchmarking Participants						
Abu Dhabi, UAE	3,980	431.498	15.881	0.799	16.680	4.084
Dubai, UAE	7,471	527.726	2.015	0.505	2.520	1.588

Summary Statistics and Standard Errors for Proficiency in ePIRLS Retrieving and Straightforward Inferencing

Country	Sample Size	ePIRLS Retrieving and Straightforward Inferencing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Canada	8,871	540.699	8.456	0.521	8.976	2.996
Chinese Taipei	4,299	548.349	4.059	0.560	4.619	2.149
Denmark	2,506	560.208	4.383	0.644	5.027	2.242
Georgia	5,557	484.799	10.799	0.365	11.164	3.341
Ireland	2,473	565.507	5.567	0.072	5.639	2.375
Israel	3,798	536.308	4.781	1.444	6.225	2.495
Italy	3,767	534.097	4.423	0.117	4.540	2.131
Norway (5)	3,610	567.395	4.441	0.470	4.911	2.216
Portugal	4,558	524.631	4.412	1.354	5.766	2.401
Singapore	6,320	594.394	10.090	0.597	10.687	3.269
Slovenia	4,303	525.401	3.086	0.194	3.280	1.811
Sweden	3,879	560.546	4.898	0.094	4.992	2.234
United Arab Emirates	15,566	470.775	4.320	0.286	4.606	2.146
United States	4,090	553.151	6.374	0.194	6.568	2.563
Benchmarking Participants						
Abu Dhabi, UAE	3,980	434.005	16.241	0.383	16.624	4.077
Dubai, UAE	7,471	527.813	2.229	0.809	3.038	1.743

Summary Statistics and Standard Errors for Proficiency in ePIRLS Interpreting, Integrating, and Evaluating

Country	Sample Size	ePIRLS Interpreting, Integrating, and Evaluating				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Canada	8,871	544.917	9.224	0.975	10.199	3.194
Chinese Taipei	4,299	543.919	3.423	0.052	3.475	1.864
Denmark	2,506	556.035	4.895	1.771	6.666	2.582
Georgia	5,557	465.986	11.288	2.513	13.800	3.715
Ireland	2,473	568.223	6.160	0.229	6.389	2.528
Israel	3,798	534.898	5.525	0.476	6.001	2.450
Italy	3,767	530.876	4.461	0.635	5.096	2.257
Norway (5)	3,610	567.564	4.686	0.452	5.138	2.267
Portugal	4,558	520.683	3.814	0.755	4.569	2.138
Singapore	6,320	584.729	8.428	1.119	9.547	3.090
Slovenia	4,303	523.398	3.737	0.142	3.879	1.970
Sweden	3,879	558.992	5.436	0.853	6.289	2.508
United Arab Emirates	15,566	465.079	4.230	0.430	4.660	2.159
United States	4,090	559.857	6.752	0.165	6.918	2.630
Benchmarking Participants						
Abu Dhabi, UAE	3,980	428.397	15.538	0.543	16.081	4.010
Dubai, UAE	7,471	527.349	1.724	0.883	2.607	1.615