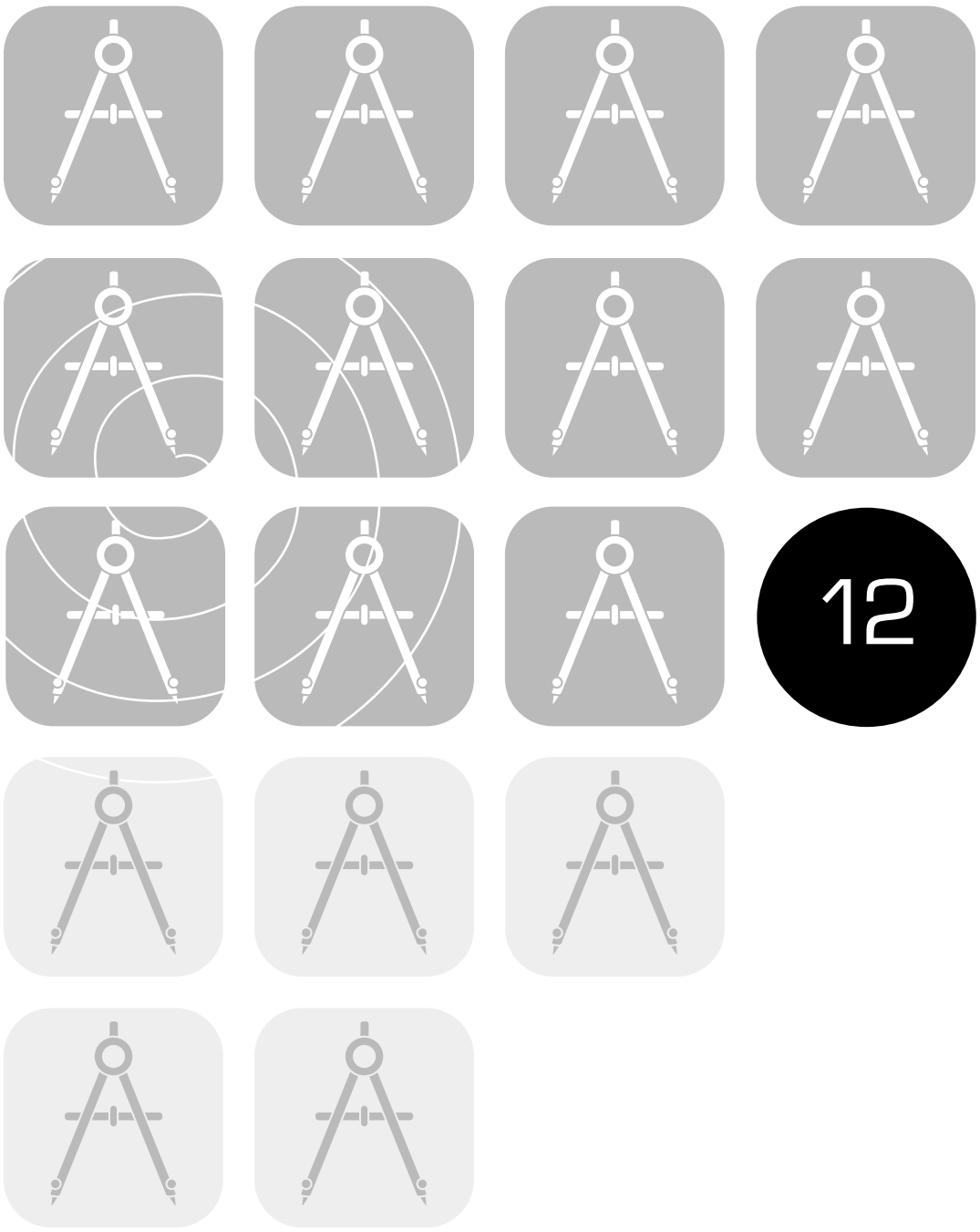
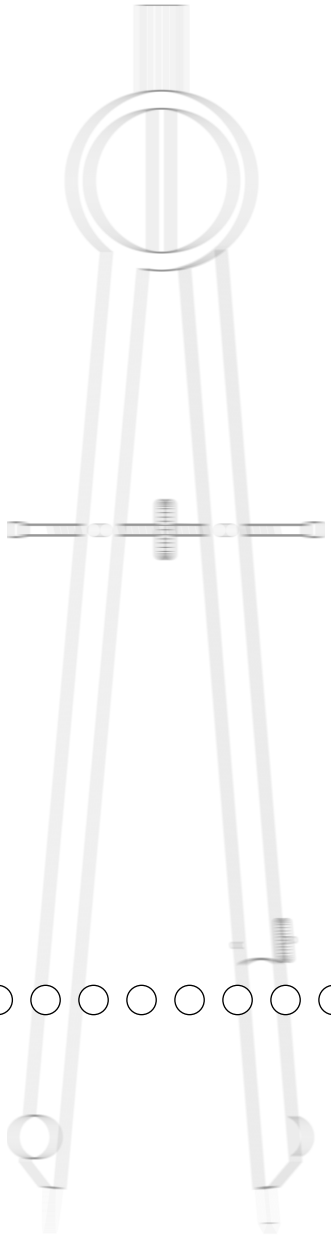


Estimation of Sampling Variance

Eugenio J. Gonzalez
Pierre Foy



12 Estimation of Sampling Variance

Eugenio J. Gonzalez
Pierre Foy

12.1 Overview

To obtain estimates of student proficiency in mathematics and science that were both accurate and cost-effective, TIMSS 1999 made extensive use of probability sampling techniques to sample students from national student populations.[1] Statistics computed from these national probability samples were used as estimates of population parameters. Because some uncertainty is involved in generalizing from samples to populations, the important statistics in the TIMSS 1999 international reports (Mullis et al., 2000; Martin et al., 2000) are presented together with their standard errors, which are a measure of this uncertainty.

The TIMSS 1999 item pool was far too extensive to be administered in its entirety to any one student, and so a complex test design was developed whereby each student was given a single test booklet containing only a part of the entire assessment.[2] The results for all of the booklets were then aggregated using item response theory to provide results for the entire assessment. Thus each student responded to just a few items from each content area, and therefore multiple imputation or "plausible values" had to be used to derive reliable indicators of student proficiency. Since every proficiency estimate incorporates some uncertainty, TIMSS followed the customary procedure of generating five estimates for each student and using the variability among them as a measure of this imputation uncertainty, or error. In the TIMSS 1999 international report the imputation error for each variable has been combined with the sampling error for that variable to provide a standard error incorporating both.

12.2 Estimating Sampling Variance

The TIMSS 1999 sampling design applied a stratified multistage cluster-sampling technique to the problem of selecting efficient and accurate samples of students while working with schools and classes. This design capitalized on the structure of the student population (i.e., students grouped in classes within schools) to

○○○

1. The TIMSS sample design is presented in Chapter 2.
2. Details of the TIMSS test design may be found in Chapter 3.

derive student samples that permitted efficient and economical data collection. Unfortunately, however, such a complex sampling design complicated the task of computing standard errors to quantify sampling variability.

When, as in TIMSS, the sampling design involves multistage cluster sampling, there are several options for estimating sampling errors that avoid the assumption of simple random sampling (Wolter, 1985). The jackknife repeated replication technique (JRR) was chosen by TIMSS in both 1995 and 1999 because it is computationally straightforward and provides approximately unbiased estimates of the sampling errors of means, totals, and percentages.

The variation on the JRR technique used in TIMSS 1999 is described in Johnson and Rust (1992). It assumes that the primary sampling units (PSUs) can be paired in a manner consistent with the sample design, with each pair regarded as members of a pseudo-stratum for variance estimation purposes. When used in this way, the JRR technique appropriately accounts for the combined effect of the between- and within-PSU contributions to the sampling variance. The general use of JRR entails systematically assigning pairs of schools to sampling zones, and randomly selecting one of these schools to have its contribution doubled and the other to have it zeroed, so as to construct a number of "pseudo-replicates" of the original sample. The statistic of interest is computed once for all of the original sample, and once again for each pseudo-replicate sample. The variation between the estimates for each of the replicate samples and the original sample estimate is the jackknife estimate of the sampling error of the statistic.

12.2.1 Construction of Sampling Zones for Sampling Variance Estimation

To apply the JRR technique used in TIMSS 1999 the sampled schools had to be 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 range of sampling zones used in each country.

Exhibit 12.1 Range of Sampling Zones Used in Each Country

Country	Zones	Country	Zones
Australia	75	Latvia (LSS)	73
Belgium (Flemish)	74	Lithuania	75
Bulgaria	75	Macedonia, Rep. of	75
Canada	75	Malaysia	75
Chile	75	Moldova	75
Chinese Taipei	75	Morocco	75
Cyprus	61	Netherlands	63
Czech Republic	71	New Zealand	75
England	64	Philippines	75
Finland	75	Romania	74
Hong Kong, SAR	69	Russian Federation	56
Hungary	74	Singapore	73
Indonesia	75	Slovak Republic	73
Iran, Islamic Rep.	75	Slovenia	75
Israel	70	South Africa	75
Italy	75	Thailand	75
Japan	71	Tunisia	75
Jordan	74	Turkey	62
Korea, Rep. of	75	United States	53

12.2.2 Computing Sampling Variance Using the JRR Method

The JRR algorithm used in TIMSS 1999 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:

$$Var_{jrr}(t) = \sum_{h=1}^{H} [t(J_h) - t(S)]^2$$

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 $t(J_h)$ denotes the same statistic using the h^{th} jackknife replicate. This is computed using all cases except those in the h^{th} zone of the sample; for those in the h^{th} zone, all cases associated with one of the randomly selected units of the pair are removed, and the elements associated with the other unit in the zone are included twice. In practice, this is effectively accomplished by recoding to zero the weights for the cases of the element of the pair to be excluded from the replication, and multiplying by two the weights of the remaining element within the h^{th} pair.

The computation of the JRR variance estimate for any statistic in TIMSS 1999 required the computation of the statistic up to 76 times for any given country: once to obtain the statistic for the full sample, and up to 75 times to obtain the statistics for each of the jackknife replicates (J_h). The number of times a statistic needed to be computed for a given country depended on the number of implicit strata or sampling zones defined for that country.

Doubling and zeroing the weights of the selected units within the sampling zones was accomplished effectively by creating replicate weights that were then used in the calculations. This approach requires the user to temporarily create a new set of weights for each pseudo-replicate sample. Each replicate weight is equal to k times the overall sampling weight, where k can take values of 0, 1, or 2 depending on whether the case is to be removed from the computation, left as it is, or have its weight doubled. The value of k for an individual student record for a given replicate depends on the assignment of the record to the specific PSU and zone.

Within each zone the members of the pair of schools are assigned an indicator (u_i), coded randomly to 1 or 0 so that one of them has a value of 1 on the variable u_i, and the other a value of 0. This indicator determines whether the weights for the elements in the school in this zone are to be doubled or zeroed. The replicate weight $(W_h^{g,i,j})$ for the elements in a school assigned to zone h is computed as the product of k_h times their overall sampling weight, where k_h can take values of 0, 1, or 2 depending on

whether the school is to be omitted, be included with its usual weight, or have its weight doubled for the computation of the statistic of interest. In TIMSS 1999, the replicate 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, $W_h^{g,i,j}$, where h takes values from 1 to 75. The value of $W_0^{g,i,j}$, 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 11.

The replicate weights for a single case were then computed as

$$W_h^{g,i,j} = W_0^{g,i,j} \cdot k_{hi}$$

where the variable k_h for an individual i takes the value $k_{hi} = 2 * u_i$ if the record belongs to zone h, and $k_{hi} = 1$ otherwise.

In the TIMSS 1999 analysis, 75 replicate weights were computed for each country regardless of the number of actual zones within the country. If a country had fewer than 75 zones, then the replicate weights W_h, where h was greater than the number of zones within the country, were each the same as the overall sampling weight. Although this involved some redundant computation, having 75 replicate weights for each country had no effect on the size of the error variance computed using the jackknife formula, but it facilitated the computation of standard errors for a number of countries at a time.

Although standard errors presented in the international reports were computed using SAS programs developed at the International Study Center, they were also verified against results produced by the WesVarPC software (Westat, 1997) as an additional quality control check.

12.3 Estimating Imputation Variance

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 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, where m = 1, 2, …, 5.

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.

12.4 Combining Sampling and Imputation Variance

When reporting standard errors for proficiency estimates using plausible values, it was necessary to combine the sampling and imputation components of the error variance for the estimate. Under ideal circumstances and with unlimited computing resources, the user would compute the imputation variance for the plausible values and the JRR sampling variance for each of the plausible values. This would be equivalent to computing the same statistic up to 380 times (once overall for each of the five plausible values using the overall sampling weights, and then 75 times more for each plausible value using the complete set of replicate weights). An acceptable shortcut, however, is to compute the JRR variance component using one plausible value, and then the imputation variance using the five plausible values. Using this approach, the same statistic needed to be computed only 80 times. With this procedure the error variance component for a statistic was computed using the following formula: $Var \cdot (t_{pv}) = Var_{jrr}(t_1) + Var_{imp}$

where $Var_{jrr}(t_1)$ is the sampling variance for the first plausible value. The User Guide for the TIMSS 1999 International Database will contain programs in SAS and SPSS that compute each of these variance components for the TIMSS 1999 data.

Exhibits 12.2 through 12.14 show basic summary statistics for mathematics and its five content areas: algebra; data representation, analysis and probability; fractions and number sense; geometry; and measurement, and for science and its six content areas: chemistry; earth science; environment and resource issues; life science; physics; and scientific inquiry and the nature of science. Each exhibit presents the student sample size, the mean and standard deviation, averaged across the five plausible values, the jackknife standard error for the mean, and the overall standard errors for the mean including imputation error.

Exhibit 12.2 Summary Statistics and Standard Errors for Mathematics Proficiency

Country	Sample Size	Mean Proficiency[a]	Standard Deviation[a]	Jackknife Sampling Error	Overall Standard Error[b]
Australia	4032	525	80	4.7	4.8
Belgium (Flemish)	5259	558	77	3.1	3.3
Bulgaria	3272	511	86	5.8	5.8
Canada	8770	531	73	2.2	2.5
Chile	5907	392	85	4.1	4.4
Chinese Taipei	5772	585	104	3.9	4.0
Cyprus	3116	476	82	1.6	1.8
Czech Republic	3453	520	79	4.1	4.2
England	2960	496	83	4.1	4.1
Finland	2920	520	65	2.6	2.7
Hong Kong, SAR	5179	582	73	4.2	4.3
Hungary	3183	532	85	3.6	3.7
Indonesia	5848	403	101	4.6	4.9
Iran, Islamic Rep.	5301	422	83	3.2	3.4
Israel	4195	466	96	3.9	3.9
Italy	3328	479	87	3.8	3.8
Japan	4745	579	80	1.5	1.7
Jordan	5052	428	103	3.4	3.6
Korea, Rep. of	6114	587	79	1.7	2.0
Latvia (LSS)	2873	505	78	3.3	3.4
Lithuania	2361	482	78	4.0	4.3
Macedonia, Rep. of	4023	447	93	4.2	4.2
Malaysia	5577	519	81	4.3	4.4
Moldova	3711	469	85	3.8	3.9
Morocco	5402	337	91	1.8	2.6
Netherlands	2962	540	73	6.9	7.1
New Zealand	3613	491	89	5.1	5.2
Philippines	6601	345	97	5.5	6.0
Romania	3425	472	93	5.6	5.8
Russian Federation	4332	526	86	5.9	5.9
Singapore	4966	604	79	6.1	6.3
Slovak Republic	3497	534	75	3.9	4.0
Slovenia	3109	530	83	2.7	2.8
South Africa	8146	275	109	5.8	6.8
Thailand	5732	467	85	4.8	5.1
Tunisia	5051	448	64	2.1	2.4
Turkey	7841	429	86	4.0	4.3
United States	9072	502	88	3.9	4.0

a. Average across the five plausible values.

b. Includes error due to sampling and imputation.

Exhibit 12.3 Summary Statistics and Standard Errors for Geometry Proficiency

Country	Sample Size	Mean Proficiency[a]	Standard Deviation[a]	Jackknife Sampling Error	Overall Standard Error[b]
Australia	4032	497	91	3.5	5.7
Belgium (Flemish)	5259	535	101	3.1	4.1
Bulgaria	3272	524	107	4.8	5.9
Canada	8770	507	89	1.5	4.7
Chile	5907	412	102	3.3	5.4
Chinese Taipei	5772	557	104	3.2	5.8
Cyprus	3116	484	90	2.0	4.6
Czech Republic	3453	513	107	3.8	5.5
England	2960	471	86	3.0	4.2
Finland	2920	494	100	3.3	6.0
Hong Kong, SAR	5179	556	88	3.3	4.9
Hungary	3183	489	108	3.5	4.3
Indonesia	5848	441	103	3.7	5.1
Iran, Islamic Rep.	5301	447	93	2.7	2.9
Israel	4195	462	102	4.1	5.4
Italy	3328	482	96	3.0	5.6
Japan	4745	575	98	2.5	5.1
Jordan	5052	449	101	2.6	7.1
Korea, Rep. of	6114	573	98	2.0	3.9
Latvia (LSS)	2873	522	94	2.5	5.6
Lithuania	2361	496	95	3.7	5.8
Macedonia, Rep. of	4023	460	114	3.5	6.1
Malaysia	5577	497	93	3.7	4.4
Moldova	3711	481	112	3.6	5.0
Morocco	5402	407	113	1.9	2.2
Netherlands	2962	515	92	4.9	5.5
New Zealand	3613	478	86	3.6	4.2
Philippines	6601	383	93	3.0	3.4
Romania	3425	487	111	3.9	6.4
Russian Federation	4332	522	113	4.7	6.0
Singapore	4966	560	93	4.9	6.7
Slovak Republic	3497	527	91	3.5	7.3
Slovenia	3109	506	111	3.1	6.2
South Africa	8146	335	106	3.8	6.6
Thailand	5732	484	90	2.8	4.4
Tunisia	5051	484	83	1.7	4.4
Turkey	7841	428	101	4.3	5.7
United States	9072	473	90	2.3	4.4

a. Average across the five plausible values.
b. Includes error due to sampling and imputation.

Exhibit 12.4 Summary Statistics and Standard Errors for Data Representation, Analysis and Probability Proficiency

Country	Sample Size	Mean Proficiency[a]	Standard Deviation[a]	Jackknife Sampling Error	Overall Standard Error[b]
Australia	4032	522	97	4.5	6.3
Belgium (Flemish)	5259	544	103	3.7	3.8
Bulgaria	3272	493	112	5.3	6.1
Canada	8770	521	93	2.5	4.5
Chile	5907	429	90	3.0	3.8
Chinese Taipei	5772	559	108	3.2	5.1
Cyprus	3116	472	94	1.5	4.6
Czech Republic	3453	513	107	3.8	5.9
England	2960	506	94	4.3	8.0
Finland	2920	525	105	2.9	3.8
Hong Kong, SAR	5179	547	89	3.7	5.4
Hungary	3183	520	118	3.9	5.9
Indonesia	5848	423	93	3.1	4.4
Iran, Islamic Rep.	5301	430	89	2.9	6.0
Israel	4195	468	102	3.9	5.1
Italy	3328	484	101	3.8	4.5
Japan	4745	555	89	2.0	2.3
Jordan	5052	436	98	2.5	7.8
Korea, Rep. of	6114	576	98	1.7	4.2
Latvia (LSS)	2873	495	104	3.2	4.8
Lithuania	2361	493	88	3.2	3.6
Macedonia, Rep. of	4023	442	111	3.7	6.2
Malaysia	5577	491	86	3.2	4.0
Moldova	3711	450	104	3.1	5.7
Morocco	5402	383	101	1.8	3.5
Netherlands	2962	538	98	7.1	7.9
New Zealand	3613	497	97	4.5	5.0
Philippines	6601	406	82	2.5	3.5
Romania	3425	453	110	3.8	4.7
Russian Federation	4332	501	110	4.5	4.8
Singapore	4966	562	94	5.6	6.2
Slovak Republic	3497	521	101	4.0	4.6
Slovenia	3109	530	114	2.8	4.2
South Africa	8146	356	94	3.3	3.8
Thailand	5732	476	91	3.6	4.0
Tunisia	5051	446	79	1.6	5.1
Turkey	7841	446	87	2.9	3.3
United States	9072	506	102	3.7	5.2

a. Average across the five plausible values.
b. Includes error due to sampling and imputation.

Exhibit 12.5 Summary Statistics and Standard Errors for Measurement Proficiency

Country	Sample Size	Mean Proficiency[a]	Standard Deviation[a]	Jackknife Sampling Error	Overall Standard Error[b]
Australia	4032	529	84	3.8	4.9
Belgium (Flemish)	5259	549	77	2.9	4.0
Bulgaria	3272	497	96	5.4	6.6
Canada	8770	521	80	2.0	2.4
Chile	5907	412	92	3.3	4.9
Chinese Taipei	5772	566	96	3.1	3.4
Cyprus	3116	471	93	2.2	4.0
Czech Republic	3453	535	83	3.3	5.0
England	2960	507	84	3.7	3.8
Finland	2920	521	74	2.6	4.7
Hong Kong, SAR	5179	567	79	4.0	5.8
Hungary	3183	538	84	2.6	3.5
Indonesia	5848	395	117	4.4	5.1
Iran, Islamic Rep.	5301	401	100	3.5	4.7
Israel	4195	457	97	3.9	5.1
Italy	3328	501	89	3.4	5.0
Japan	4745	558	75	1.7	2.4
Jordan	5052	438	106	3.2	4.4
Korea, Rep. of	6114	571	79	1.9	2.8
Latvia (LSS)	2873	505	89	3.1	3.5
Lithuania	2361	467	81	3.1	4.0
Macedonia, Rep. of	4023	451	101	3.4	5.2
Malaysia	5577	514	86	4.1	4.6
Moldova	3711	479	97	3.5	4.9
Morocco	5402	348	115	2.2	3.5
Netherlands	2962	538	73	5.4	5.8
New Zealand	3613	496	86	4.4	5.3
Philippines	6601	355	104	4.2	6.2
Romania	3425	491	99	4.4	4.9
Russian Federation	4332	527	94	5.5	6.0
Singapore	4966	599	87	5.6	6.3
Slovak Republic	3497	537	77	3.0	3.3
Slovenia	3109	523	94	2.7	3.7
South Africa	8146	329	108	3.7	4.8
Thailand	5732	463	92	4.4	6.2
Tunisia	5051	442	81	2.3	3.1
Turkey	7841	436	93	4.5	6.5
United States	9072	482	92	3.5	3.9

a. Average across the five plausible values.
b. Includes error due to sampling and imputation.

Exhibit 12.6 Summary Statistics and Standard Errors for Algebra Proficiency

Country	Sample Size	Mean Proficiency[a]	Standard Deviation[a]	Jackknife Sampling Error	Overall Standard Error[b]
Australia	4032	520	81	4.1	5.1
Belgium (Flemish)	5259	540	86	3.2	4.6
Bulgaria	3272	512	88	4.8	5.1
Canada	8770	525	73	1.7	2.4
Chile	5907	399	96	3.9	4.3
Chinese Taipei	5772	586	114	4.3	4.4
Cyprus	3116	479	80	1.5	1.6
Czech Republic	3453	514	87	3.8	4.0
England	2960	498	77	3.3	4.9
Finland	2920	498	73	2.3	3.1
Hong Kong, SAR	5179	569	78	3.6	4.5
Hungary	3183	536	94	3.4	4.1
Indonesia	5848	424	104	3.9	5.7
Iran, Islamic Rep.	5301	434	88	2.8	4.9
Israel	4195	479	97	4.1	4.5
Italy	3328	481	84	3.3	3.6
Japan	4745	569	82	1.5	3.3
Jordan	5052	439	108	3.6	5.3
Korea, Rep. of	6114	585	90	1.9	2.7
Latvia (LSS)	2873	499	83	3.0	4.3
Lithuania	2361	487	74	3.4	3.7
Macedonia, Rep. of	4023	465	100	3.8	4.0
Malaysia	5577	505	81	3.8	4.8
Moldova	3711	477	91	3.2	3.7
Morocco	5402	353	111	2.2	4.7
Netherlands	2962	522	77	6.9	7.7
New Zealand	3613	497	81	4.3	4.7
Philippines	6601	345	119	5.2	5.8
Romania	3425	481	99	5.0	5.2
Russian Federation	4332	529	95	4.8	4.9
Singapore	4966	576	81	5.9	6.2
Slovak Republic	3497	525	76	3.6	4.6
Slovenia	3109	525	85	2.7	2.9
South Africa	8146	293	125	6.1	7.7
Thailand	5732	456	91	4.2	4.9
Tunisia	5051	455	74	1.9	2.7
Turkey	7841	432	98	4.3	4.6
United States	9072	506	90	3.4	4.1

a. Average across the five plausible values.

b. Includes error due to sampling and imputation.

Exhibit 12.7 Summary Statistics and Standard Errors for Fractions and Number Sense Proficiency

Country	Sample Size	Mean Proficiency[a]	Standard Deviation[a]	Jackknife Sampling Error	Overall Standard Error[b]
Australia	4032	519	78	4.1	4.3
Belgium (Flemish)	5259	557	74	2.8	3.1
Bulgaria	3272	503	97	6.3	6.6
Canada	8770	533	74	1.9	2.5
Chile	5907	403	88	3.6	4.9
Chinese Taipei	5772	576	101	3.8	4.2
Cyprus	3116	481	82	2.0	3.0
Czech Republic	3453	507	90	4.0	4.8
England	2960	497	82	3.7	3.8
Finland	2920	531	75	3.1	3.8
Hong Kong, SAR	5179	579	75	4.0	4.5
Hungary	3183	526	95	3.8	4.2
Indonesia	5848	406	99	3.9	4.1
Iran, Islamic Rep.	5301	437	82	2.8	4.5
Israel	4195	472	93	4.0	4.4
Italy	3328	471	88	3.6	5.0
Japan	4745	570	84	1.6	2.6
Jordan	5052	432	101	2.9	3.2
Korea, Rep. of	6114	570	78	1.9	2.7
Latvia (LSS)	2873	496	89	3.6	3.7
Lithuania	2361	479	84	4.0	4.3
Macedonia, Rep. of	4023	437	100	4.1	4.7
Malaysia	5577	532	83	4.2	4.7
Moldova	3711	465	92	3.7	4.2
Morocco	5402	335	113	1.8	3.6
Netherlands	2962	545	79	6.7	7.1
New Zealand	3613	493	88	4.5	5.0
Philippines	6601	378	97	4.7	6.3
Romania	3425	458	100	5.3	5.7
Russian Federation	4332	513	98	6.1	6.4
Singapore	4966	608	82	5.4	5.6
Slovak Republic	3497	525	81	4.6	4.8
Slovenia	3109	527	90	3.1	3.7
South Africa	8146	300	115	5.2	6.0
Thailand	5732	471	90	4.4	5.3
Tunisia	5051	443	79	2.2	2.8
Turkey	7841	430	88	3.6	4.3
United States	9072	509	88	3.8	4.2

a. Average across the five plausible values.
b. Includes error due to sampling and imputation.

Exhibit 12.8 Summary Statistics and Standard Errors for Science Proficiency

Country	Sample Size	Mean of 5 Plausible Values	S.D.[a]	Error Due to Sampling	S.E.[b]
Australia	4032	540	87	4.3	4.4
Belgium (Flemish)	5259	535	69	2.6	3.1
Bulgaria	3272	518	93	5.3	5.4
Canada	8770	533	78	1.8	2.1
Chile	5907	420	88	3.7	3.7
Chinese Taipei	5772	569	89	3.6	4.4
Cyprus	3116	460	84	1.8	2.4
Czech Republic	3453	539	80	3.7	4.2
England	2960	538	91	4.3	4.8
Finland	2920	535	78	3.0	3.5
Hong Kong, SAR	5179	530	70	3.5	3.7
Hungary	3183	552	84	3.4	3.7
Indonesia	5848	435	84	4.1	4.5
Iran, Islamic Rep.	5301	448	84	3.7	3.8
Israel	4195	468	105	4.4	4.9
Italy	3328	493	87	3.5	3.9
Japan	4745	550	76	1.9	2.2
Jordan	5052	450	103	3.4	3.8
Korea, Rep. of	6114	549	85	1.9	2.6
Latvia (LSS)	2873	503	78	3.1	4.8
Lithuania	2361	488	83	3.8	4.1
Macedonia, Rep. of	4023	458	97	4.3	5.2
Malaysia	5577	492	82	4.2	4.4
Moldova	3711	459	95	3.9	4.0
Morocco	5402	323	102	2.9	4.3
Netherlands	2962	545	77	6.7	6.9
New Zealand	3613	510	93	4.6	4.9
Philippines	6601	345	121	7.2	7.5
Romania	3425	472	97	5.0	5.8
Russian Federation	4332	529	93	6.1	6.4
Singapore	4966	568	97	8.0	8.0
Slovak Republic	3497	535	78	3.0	3.3
Slovenia	3109	533	84	2.9	3.2
South Africa	8146	243	132	7.4	7.8
Thailand	5732	482	73	3.9	4.0
Tunisia	5051	430	67	2.0	3.4
Turkey	7841	433	80	3.5	4.3
United States	9072	515	97	4.4	4.6

a. Standard deviation of the five plausible values
b. Standard error due to imputation

Exhibit 12.9 Summary Statistics and Standard Errors for Life Science Proficiency Sample

Country	Sample Size	Mean of 5 Plausible Values	S.D.[a]	Error Due to Sampling	S.E.[b]
Australia	4032	530	96	4.0	4.4
Belgium (Flemish)	5259	535	89	2.8	4.6
Bulgaria	3272	514	107	5.4	6.9
Canada	8770	523	87	2.1	3.8
Chile	5907	431	88	3.0	3.7
Chinese Taipei	5772	550	96	2.8	3.3
Cyprus	3116	468	94	2.1	3.8
Czech Republic	3453	544	99	3.7	4.1
England	2960	533	97	4.3	6.2
Finland	2920	520	94	2.5	4.0
Hong Kong, SAR	5179	516	84	3.1	5.5
Hungary	3183	535	99	3.3	4.0
Indonesia	5848	448	85	3.1	3.6
Iran, Islamic Rep.	5301	437	92	2.7	3.7
Israel	4195	463	103	3.8	4.0
Italy	3328	488	94	3.3	4.6
Japan	4745	534	90	2.1	5.4
Jordan	5052	448	103	3.3	4.1
Korea, Rep. of	6114	528	93	2.0	3.6
Latvia (LSS)	2873	509	90	3.1	3.9
Lithuania	2361	494	87	3.5	4.6
Macedonia, Rep. of	4023	468	113	4.0	4.9
Malaysia	5577	479	94	4.1	5.4
Moldova	3711	477	109	3.7	3.9
Morocco	5402	347	108	1.9	2.8
Netherlands	2962	536	94	6.0	7.2
New Zealand	3613	501	98	4.5	5.6
Philippines	6601	378	110	5.6	5.7
Romania	3425	475	109	4.7	6.0
Russian Federation	4332	517	114	5.7	6.5
Singapore	4966	541	102	7.1	7.2
Slovak Republic	3497	535	93	3.6	6.2
Slovenia	3109	521	103	2.8	3.9
South Africa	8146	289	123	6.2	7.3
Thailand	5732	508	77	2.7	4.5
Tunisia	5051	441	76	1.7	5.0
Turkey	7841	444	85	3.7	4.5
United States	9072	520	104	3.7	4.1

a. Standard deviation of the five plausible values

b. Standard error due to imputation

Exhibit 12.10 Summary Statistics and Standard Errors for Earth Science Proficiency

Country	Sample Size	Mean Proficiency[a]	Standard Deviation[a]	Jackknife Sampling Error	Overall Standard Error[b]
Australia	4032	519	96	3.9	6.1
Belgium (Flemish)	5259	533	92	2.8	3.5
Bulgaria	3272	520	115	5.4	5.7
Canada	8770	519	92	1.7	3.7
Chile	5907	435	93	3.0	7.0
Chinese Taipei	5772	538	89	2.0	3.0
Cyprus	3116	459	87	1.8	5.4
Czech Republic	3453	533	113	4.7	6.9
England	2960	525	88	3.6	3.9
Finland	2920	520	101	3.0	5.5
Hong Kong, SAR	5179	506	82	2.5	4.3
Hungary	3183	560	119	3.8	3.9
Indonesia	5848	431	99	3.7	6.4
Iran, Islamic Rep.	5301	459	96	2.8	5.2
Israel	4195	472	108	4.4	5.2
Italy	3328	502	103	3.6	5.9
Japan	4745	533	91	2.2	6.2
Jordan	5052	446	92	2.4	3.5
Korea, Rep. of	6114	532	98	2.1	2.7
Latvia (LSS)	2873	495	114	3.8	5.4
Lithuania	2361	476	91	3.2	4.4
Macedonia, Rep. of	4023	464	116	3.9	4.2
Malaysia	5577	491	90	3.4	4.2
Moldova	3711	466	117	3.0	4.2
Morocco	5402	363	112	2.0	3.3
Netherlands	2962	534	94	6.0	7.2
New Zealand	3613	504	90	3.7	5.8
Philippines	6601	390	103	4.9	5.0
Romania	3425	475	128	4.5	5.5
Russian Federation	4332	529	124	4.5	5.1
Singapore	4966	521	91	5.4	7.3
Slovak Republic	3497	537	99	4.0	4.3
Slovenia	3109	541	111	3.6	4.3
South Africa	8146	348	102	3.6	4.8
Thailand	5732	470	95	3.4	3.9
Tunisia	5051	442	89	1.6	2.7
Turkey	7841	435	90	3.6	4.6
United States	9072	504	98	3.4	4.2

a. Average across the five plausible values.

b. Includes error due to sampling and imputation.

Exhibit 12.11 Summary Statistics and Standard Errors for Physics Proficiency

Country	Sample Size	Mean Proficiency[a]	Standard Deviation[a]	Jackknife Sampling Error	Overall Standard Error[b]
Australia	4032	531	90	3.6	6.3
Belgium (Flemish)	5259	530	82	2.0	3.5
Bulgaria	3272	505	109	4.8	5.8
Canada	8770	521	85	2.3	3.8
Chile	5907	428	93	2.6	5.6
Chinese Taipei	5772	552	96	3.0	3.9
Cyprus	3116	459	95	2.0	2.9
Czech Republic	3453	526	99	3.6	4.2
England	2960	528	86	3.7	4.5
Finland	2920	520	103	2.6	4.4
Hong Kong, SAR	5179	523	88	3.4	4.9
Hungary	3183	543	102	3.0	4.3
Indonesia	5848	452	94	3.2	5.5
Iran, Islamic Rep.	5301	445	105	4.0	5.7
Israel	4195	484	102	3.9	5.3
Italy	3328	480	93	3.5	4.1
Japan	4745	544	83	1.7	2.9
Jordan	5052	459	108	3.1	3.6
Korea, Rep. of	6114	544	92	2.3	5.1
Latvia (LSS)	2873	495	95	3.1	3.9
Lithuania	2361	510	85	3.5	4.3
Macedonia, Rep. of	4023	463	107	3.8	6.0
Malaysia	5577	494	89	3.2	4.1
Moldova	3711	457	112	3.9	5.5
Morocco	5402	352	120	2.2	4.2
Netherlands	2962	537	91	6.5	6.5
New Zealand	3613	499	93	3.7	4.7
Philippines	6601	393	107	5.1	6.3
Romania	3425	465	110	4.4	6.8
Russian Federation	4332	529	115	5.9	6.3
Singapore	4966	570	96	6.4	6.7
Slovak Republic	3497	518	91	3.5	4.1
Slovenia	3109	525	102	3.4	4.4
South Africa	8146	308	122	5.9	6.7
Thailand	5732	475	90	4.0	4.2
Tunisia	5051	425	87	2.2	6.3
Turkey	7841	441	93	3.9	4.0
United States	9072	498	97	3.7	5.5

a. Average across the five plausible values.

b. Includes error due to sampling and imputation.

Exhibit 12.12 Summary Statistics and Standard Errors for Chemistry Proficiency

Country	Sample Size	Mean Proficiency[a]	Standard Deviation[a]	Jackknife Sampling Error	Overall Standard Error[b]
Australia	4032	520	101	4.2	5.0
Belgium (Flemish)	5259	508	92	2.4	3.3
Bulgaria	3272	527	115	4.5	5.7
Canada	8770	521	94	2.0	5.4
Chile	5907	435	97	3.2	5.2
Chinese Taipei	5772	563	105	3.0	4.3
Cyprus	3116	470	91	1.7	3.4
Czech Republic	3453	512	108	3.5	5.2
England	2960	524	95	3.8	5.5
Finland	2920	535	101	3.0	4.5
Hong Kong, SAR	5179	515	87	2.6	5.2
Hungary	3183	548	111	3.1	4.7
Indonesia	5848	425	88	3.5	3.9
Iran, Islamic Rep.	5301	487	92	2.4	4.1
Israel	4195	479	107	3.8	4.7
Italy	3328	493	94	3.2	4.8
Japan	4745	530	87	1.8	3.1
Jordan	5052	483	112	3.0	5.5
Korea, Rep. of	6114	523	102	2.8	3.7
Latvia (LSS)	2873	490	104	2.9	3.7
Lithuania	2361	485	95	3.8	4.6
Macedonia, Rep. of	4023	481	113	3.7	6.1
Malaysia	5577	485	91	2.9	3.5
Moldova	3711	451	117	3.7	5.6
Morocco	5402	372	107	1.7	4.8
Netherlands	2962	515	95	5.2	6.4
New Zealand	3613	503	96	3.8	4.9
Philippines	6601	394	100	4.2	6.5
Romania	3425	481	115	4.1	6.1
Russian Federation	4332	523	120	6.8	8.0
Singapore	4966	545	116	7.9	8.3
Slovak Republic	3497	525	101	3.4	4.9
Slovenia	3109	509	112	2.5	5.4
South Africa	8146	350	105	3.1	4.0
Thailand	5732	439	97	4.0	4.3
Tunisia	5051	439	83	1.7	3.7
Turkey	7841	437	98	3.1	5.0
United States	9072	508	110	4.0	4.8

a. Average across the five plausible values.

b. Includes error due to sampling and imputation.

Exhibit 12.13 Summary Statistics and Standard Errors for Scientific Inquiry and the Nature of Science Proficiency

Country	Sample Size	Mean Proficiency[a]	Standard Deviation[a]	Jackknife Sampling Error	Overall Standard Error[b]
Australia	4032	535	93	3.5	4.9
Belgium (Flemish)	5259	526	93	2.7	4.9
Bulgaria	3272	479	121	5.4	5.6
Canada	8770	532	86	1.2	5.1
Chile	5907	441	100	3.3	4.7
Chinese Taipei	5772	540	87	3.0	4.9
Cyprus	3116	467	104	2.1	4.6
Czech Republic	3453	522	108	4.8	5.7
England	2960	538	86	3.2	5.1
Finland	2920	528	101	2.6	4.0
Hong Kong, SAR	5179	531	82	2.3	2.8
Hungary	3183	526	103	2.9	5.9
Indonesia	5848	446	99	2.7	4.3
Iran, Islamic Rep.	5301	446	94	2.3	5.3
Israel	4195	476	112	3.8	8.3
Italy	3328	489	96	2.9	4.6
Japan	4745	543	77	1.8	2.8
Jordan	5052	440	109	2.6	5.5
Korea, Rep. of	6114	545	89	2.1	7.3
Latvia (LSS)	2873	495	104	3.2	4.7
Lithuania	2361	483	99	4.0	6.4
Macedonia, Rep. of	4023	464	117	3.2	3.6
Malaysia	5577	488	84	2.5	4.5
Moldova	3711	471	113	3.3	3.8
Morocco	5402	391	134	2.7	4.2
Netherlands	2962	534	98	5.1	6.5
New Zealand	3613	521	95	3.3	6.8
Philippines	6601	403	108	3.7	5.5
Romania	3425	456	118	3.4	5.5
Russian Federation	4332	491	109	3.3	4.9
Singapore	4966	550	85	4.2	5.9
Slovak Republic	3497	507	85	2.7	3.9
Slovenia	3109	513	107	2.9	4.3
South Africa	8146	329	133	4.8	6.4
Thailand	5732	462	99	3.4	4.2
Tunisia	5051	451	95	2.1	3.4
Turkey	7841	445	104	4.0	6.3
United States	9072	522	92	2.6	4.3

a. Average across the five plausible values.
b. Includes error due to sampling and imputation.

Exhibit 12.14 Summary Statistics and Standard Errors for Environment and Resources Issues Proficiency

Country	Sample Size	Mean Proficiency[a]	Standard Deviation[a]	Jackknife Sampling Error	Overall Standard Error[b]
Australia	4032	530	104	3.9	6.3
Belgium (Flemish)	5259	513	98	2.3	3.5
Bulgaria	3272	483	126	5.5	6.4
Canada	8770	521	97	2.5	3.5
Chile	5907	449	97	2.6	4.8
Chinese Taipei	5772	567	101	2.4	4.0
Cyprus	3116	475	92	2.2	4.3
Czech Republic	3453	516	111	3.5	5.7
England	2960	518	108	4.1	5.8
Finland	2920	514	101	2.4	7.1
Hong Kong, SAR	5179	518	91	2.9	4.9
Hungary	3183	501	118	3.6	6.6
Indonesia	5848	489	84	2.2	4.8
Iran, Islamic Rep.	5301	470	86	2.6	5.5
Israel	4195	458	105	3.5	4.0
Italy	3328	491	93	2.5	5.4
Japan	4745	506	89	2.2	5.5
Jordan	5052	476	106	2.7	6.0
Korea, Rep. of	6114	523	96	1.5	4.5
Latvia (LSS)	2873	493	98	3.4	5.2
Lithuania	2361	458	98	3.4	5.1
Macedonia, Rep. of	4023	432	117	3.3	4.2
Malaysia	5577	502	89	3.1	4.4
Moldova	3711	444	127	3.5	6.2
Morocco	5402	396	116	3.1	5.1
Netherlands	2962	526	106	7.1	8.5
New Zealand	3613	503	99	4.4	5.2
Philippines	6601	391	114	5.8	7.6
Romania	3425	473	114	4.4	6.6
Russian Federation	4332	495	118	5.2	6.6
Singapore	4966	577	117	7.9	8.3
Slovak Republic	3497	512	94	2.8	4.5
Slovenia	3109	519	110	3.0	3.4
South Africa	8146	350	118	5.4	8.5
Thailand	5732	507	83	2.2	3.0
Tunisia	5051	462	84	1.7	5.0
Turkey	7841	461	88	2.7	3.6
United States	9072	509	107	3.6	6.4

a. Average across the five plausible values.
b. Includes error due to sampling and imputation.

References

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

Martin, M.O., Mullis, I.V.S., Gonzalez, E. J., Gregory, K.D., Smith, T.A., Chrostowski, S.J., Garden, R.A., & O'Connor, K.M. (2000). *TIMSS 1999 International Science Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: Boston College.

Mislevy, R.J., Beaton, A.E., Kaplan, B., & Sheenan, K.M. (1992). Estimating Population Characteristics from Sparse Matrix Samples of Item Responses. *Journal of Educational Measurement, 29*, 133-161.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S. J., & Smith, T.A. (2000). *TIMSS 1999 International Mathematics Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. 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.