5 Estimation of Sampling and Imputation Variability

Eugenio J. Gonzalez
Boston College

Pierre Foy
Statistics Canada

5.1 OVERVIEW

One aim of TIMSS was to obtain accurate and cost-effective estimates of student performance in the populations under study – mathematics and science literacy among students in their final year of secondary school, and advanced mathematics and physics among final-year students with preparation in these subjects. To that end, TIMSS 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 of people to populations, the important statistics in the TIMSS international report (Mullis, Martin, Beaton, Gonzalez, Kelly, and Smith, 1998) are presented together with their standard errors, which are a measure of this uncertainty.

The TIMSS 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. A consequence of this approach was that each student responded to just a few items from each content area in the assessment, and therefore multiple imputation or "plausible values" (see Chapter 7 of this volume) had to be used to derive reliable indicators of student proficiency. Since each proficiency estimate incorporates some uncertainty it is customary to generate a number of estimates (usually five) for each student, and to use the variability among the five estimates as a measure of this imputation uncertainty, or error. In the TIMSS 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.

5.2 ESTIMATING SAMPLING VARIANCE

The TIMSS sampling design applies stratified multistage cluster-sampling techniques to the problem of selecting efficient and accurate samples of students while working with schools and classes. Such complex designs capitalize on the structure of the student population (i.e., students grouped in classes within schools) to derive student

1 See Foy, Rust, and Schleicher (1996) for details of the TIMSS sampling design.

2 See Adams and Gonzalez (1996) for details of the TIMSS test design.

samples that permit efficient and economical data collection. However, complex sampling designs make the task of computing standard errors to quantify sampling variability more difficult.

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 (see Wolter, 1985). The jackknife repeated replication technique (JRR) was chosen in TIMSS because it is computationally straightforward and provides approximately unbiased estimates of the sampling errors of means, totals, and percentages in complex sample designs.

The particular variation on the JRR technique used in TIMSS is described in Johnson and Rust (1992). This method assumes that the primary sampling units (PSUs) can be paired in a manner consistent with the sample design, and each pair regarded as members of a pseudo-stratum for variance estimation purposes. Note that when using the JRR technique for the estimation of sampling variability, the approach will appropriately reflect the combined effect of the between- and within-PSU contributions to the sampling variance. The general use of the 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 from each of the replicate samples and the original sample estimate is the jackknife estimate of the sampling error of the statistic.

5.3 CONSTRUCTION OF SAMPLING ZONES FOR SAMPLING VARIANCE ESTIMATION

To apply the JRR technique used in TIMSS it is necessary to pair the sampled schools and assign them to a series of groups known as sampling zones. This is done 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. A maximum of 75 zones was used, allowing for a total of 150 schools per country. In countries where more than 150 schools were sampled, it was sometimes necessary to combine two schools before assigning them to a sampling zone.

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. Table 5.1 shows the range of sampling zones used in each country.

Table 5.1 Sampling Zones - Population 3

Country	Mathematics and Science Literacy Sample		Advanced Mathematics Sample		Physics Sample	
	First Zone	Last Zone	First Zone	Last Zone	First Zone	Last Zone
Australia	1	49	1	49	1	49
Austria	1	75	1	58	1	58
Canada	1	75	1	75	1	75
Cyprus	1	28	1	21	1	21
Czech Republic	1	75	1	45	1	45
Denmark	1	65	1	65	1	65
France	32	70	1	31	1	31
Germany	1	75	1	75	1	75
Greece	-	-	1	30	1	30
Hungary	1	75	-	-	-	-
Iceland	1	30	-	-	-	-
Italy	1	52	7	51	-	-
Latvia (LSS)	-	-	-	-	1	19
Lithuania	1	73	1	29	-	-
Netherlands	1	40	-	-	-	-
New Zealand	1	39	-	-	-	-
Norway	1	66	-	-	1	33
Russian Federation	1	41	1	41	1	41
Slovenia	1	39	1	39	1	39
South Africa	1	52	-	-	-	-
Sweden	1	75	1	75	1	75
Switzerland	1	75	20	75	20	75
United States	1	33	1	33	1	33

A dash (-) indicates the country did not participate in the assessment of this subject area.

5.4 COMPUTING SAMPLING VARIANCE USING THE JRR METHOD

The JRR algorithm used in TIMSS assumes that there are H sampling zones within each country, each one containing two sampled schools selected independently. When computing 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 hth jackknife replicate. This is computed using all cases except those in the hth zone of the sample; for those in the hth zone, all cases associated with one of the randomly selected units of the pair are removed, 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 hth pair.

The computation of the JRR variance estimate for any statistic in TIMSS requires 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 needs to be computed for a given country depends 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 is accomplished effectively by creating replicate weights that are then used in the calculations. Gonzalez, Smith, and Sibberns (1998) provide examples of how this approach allows standard statistical software such as SAS or SPSS to be used to compute JRR estimates of sampling variability in TIMSS. The replicate weight 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 the members of each pair has a value of 1 on the variable u_i, and the remaining member 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, the replicate weights are not permanent variables, but are created temporarily by the sampling variance estimation program as a useful computing device.

Replicate weights were created by the following procedure.

Each sampled student was 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 4.

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

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

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 analysis, a total of 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 has no effect on the size of the error variance computed using the jackknife formula, but facilitated the computation of standard errors for a number of countries at one time.

Example SAS and SPSS programs used to compute standard errors in TIMSS are given in Gonzalez, Smith, and Sibberns (1998). 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). Results were compared with each other for accuracy.[3]

5.5 ESTIMATING IMPUTATION VARIANCE

The general procedure for estimating the imputation variance using plausible values is as follows: 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, or 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:

$$Var_{imp} = (1 + 1/M) \times Var(t_m)$$

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.

5.6 COMBINING SAMPLING AND IMPUTATION VARIANCE

When reporting standard errors for proficiency estimates using plausible values, it is 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). However, an acceptable shortcut 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 needs to be computed only 80 times. Under this procedure the error variance component for a statistic is computed using the following formula:

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

where $Var_{jrr}(t_1)$ is the sampling variance for the first plausible value. The User Guide for the TIMSS International Database (Gonzalez, Smith, and Sibberns, 1998) contains

3 Minor differences were occasionally found between the results obtained with WesVar and those obtained with software developed in-house. However, all these differences were due to the fact that the two programs did not always choose the same PSUs in forming jackknife replicates. When identical jackknife replicates were used for both programs, the results were identical.

programs in SAS and SPSS that compute each of these variance components for the TIMSS data. Tables 5.2 through 5.14 show, for each set of plausible values, the sample size, estimate of the population size, mean of the five plausible values, error due to sampling, error due to imputation, sampling and imputation errors combined, standard deviation of the plausible values, and its standard error. These statistics are presented for males and females separately in Appendix C.

Table 5.2 Sampling and Imputation Standard Errors - Mathematics and Science Literacy Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	1941	170847	525	9.5	0.5	9.5	94.9	4.8
Austria	1962	70602	519	5.3	0.8	5.4	80.4	3.1
Canada	5232	263241	526	2.5	0.7	2.6	83.2	1.6
Cyprus	534	4556	447	2.4	0.8	2.5	73.1	2.3
Czech Republic	2167	137459	476	10.5	0.5	10.5	91.8	3.3
Denmark	2714	37872	528	3.1	0.4	3.2	81.1	2.3
France	1590	637935	505	4.8	0.8	4.9	74.0	2.7
Germany	2289	967705	496	5.1	1.6	5.4	88.9	3.2
Hungary	5091	111281	477	3.0	0.5	3.0	84.3	2.4
Iceland	1703	2308	541	1.5	0.6	1.6	77.4	1.2
Italy	1616	380834	475	5.2	0.7	5.3	83.3	4.0
Lithuania	2887	22161	465	5.7	0.5	5.8	80.4	3.3
Netherlands	1470	145916	559	4.8	1.0	4.9	84.2	4.0
New Zealand	1763	37549	525	4.6	0.7	4.7	92.0	2.4
Norway	2518	43806	536	4.0	0.5	4.0	87.8	2.1
Russian Federation	2289	1031187	476	5.8	0.8	5.8	83.1	2.9
Slovenia	1622	26644	514	8.1	0.6	8.2	82.3	4.4
South Africa	2757	374618	352	9.3	0.5	9.3	87.6	8.7
Sweden	3068	71243	555	4.3	0.5	4.3	91.3	2.2
Switzerland	3308	65140	531	5.3	1.1	5.4	87.6	2.6
United States	5807	2278258	471	3.1	0.5	3.1	89.1	2.1

S.D. = standard deviation
S.E. = standard error

Table 5.3 Sampling and Imputation Standard Errors - Mathematics Literacy Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	1941	170847	522	9.3	0.4	9.3	97.2	4.9
Austria	1962	70602	518	5.3	0.7	5.3	79.8	2.8
Canada	5232	263241	519	2.8	0.6	2.8	90.1	1.7
Cyprus	534	4556	446	2.4	0.7	2.5	72.9	2.6
Czech Republic	2167	137459	466	12.3	0.8	12.3	99.4	3.5
Denmark	2714	37872	547	3.3	0.5	3.3	86.7	2.8
France	1590	637935	523	5.0	0.8	5.1	79.2	2.8
Germany	2289	967705	495	5.6	1.9	5.9	93.7	3.2
Hungary	5091	111281	483	3.1	0.6	3.2	92.3	2.2
Iceland	1703	2308	534	2.0	0.5	2.0	87.9	1.4
Italy	1616	380834	476	5.3	1.1	5.5	87.4	3.9
Lithuania	2887	22161	469	6.0	0.7	6.1	84.6	3.5
Netherlands	1470	145916	560	4.7	0.7	4.7	90.0	3.5
New Zealand	1763	37549	522	4.4	0.8	4.5	98.2	2.2
Norway	2518	43806	528	4.1	0.5	4.1	93.9	1.9
Russian Federation	2289	1031187	471	6.1	1.0	6.2	85.5	3.2
Slovenia	1622	26644	512	8.3	0.7	8.3	86.8	4.4
South Africa	2757	374618	356	8.3	0.6	8.3	81.3	8.5
Sweden	3068	71243	552	4.3	0.5	4.3	98.7	2.3
Switzerland	3308	65140	540	5.7	1.1	5.8	88.5	2.5
United States	5807	2278258	461	3.1	0.7	3.2	91.1	1.9

S.D. = standard deviation
S.E. = standard error

Table 5.4 Sampling and Imputation Standard Errors - Science Literacy Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	1941	170847	527	9.7	1.1	9.8	99.6	5.0
Austria	1962	70602	520	5.5	1.3	5.6	87.0	3.6
Canada	5232	263241	532	2.5	0.9	2.6	84.8	1.9
Cyprus	534	4556	448	2.8	1.0	3.0	82.9	2.7
Czech Republic	2167	137459	487	8.8	0.6	8.8	91.2	3.0
Denmark	2714	37872	509	3.6	0.4	3.6	86.9	2.4
France	1590	637935	487	5.0	1.2	5.1	78.8	2.4
Germany	2289	967705	497	4.9	1.3	5.1	90.7	3.5
Hungary	5091	111281	471	3.0	0.4	3.0	86.2	2.5
Iceland	1703	2308	549	1.4	0.6	1.5	75.4	1.4
Italy	1616	380834	475	5.3	0.6	5.3	86.7	3.9
Lithuania	2887	22161	461	5.7	0.5	5.7	84.0	3.2
Netherlands	1470	145916	558	5.1	1.4	5.3	85.5	4.5
New Zealand	1763	37549	529	5.1	1.0	5.2	94.4	3.2
Norway	2518	43806	544	4.1	0.6	4.1	91.2	2.5
Russian Federation	2289	1031187	481	5.6	0.8	5.7	91.0	2.8
Slovenia	1622	26644	517	8.1	0.6	8.2	83.8	4.7
South Africa	2757	374618	349	10.4	0.8	10.5	99.6	8.7
Sweden	3068	71243	559	4.4	0.4	4.4	91.0	2.2
Switzerland	3308	65140	523	5.2	1.2	5.3	93.6	2.7
United States	5807	2278258	480	3.2	0.4	3.3	93.9	2.5

S.D. = standard deviation
S.E. = standard error

Table 5.5 Sampling and Imputation Standard Errors - Advanced Mathematics Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	645	39498	525	11.6	1.2	11.6	109.2	7.9
Austria	782	31063	436	7.2	0.6	7.2	91.3	5.5
Canada	2781	58606	509	4.2	0.8	4.3	98.3	2.4
Cyprus	391	837	518	4.1	1.5	4.3	85.4	3.0
Czech Republic	1101	19446	469	11.1	1.3	11.2	106.3	9.3
Denmark	1388	13527	522	3.3	0.8	3.4	72.9	1.9
France	1071	151531	557	3.8	0.9	3.9	70.1	2.1
Germany	2296	262789	465	5.5	0.9	5.6	85.0	3.4
Greece	456	14620	513	5.9	1.2	6.0	104.9	6.0
Italy	398	104477	474	9.5	0.6	9.6	95.2	8.1
Lithuania	734	1360	516	2.4	0.9	2.6	85.1	3.2
Russian Federation	1638	42858	542	9.2	1.1	9.2	111.6	5.6
Slovenia	1536	22881	475	9.1	0.9	9.2	93.8	3.8
Sweden	1001	16408	512	4.3	0.8	4.4	85.8	2.9
Switzerland	1404	11343	533	4.8	1.4	5.0	90.5	2.7
United States	2785	496852	442	5.8	1.1	5.9	98.0	4.1

S.D. = standard deviation
S.E. = standard error

Table 5.6 Sampling and Imputation Standard Errors Numbers, Equations and Functions Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	645	39498	517	9.2	1.4	9.4	98.0	6.3
Austria	782	31063	412	7.4	0.9	7.4	91.4	5.8
Canada	2781	58606	512	3.8	1.0	3.9	84.3	2.8
Cyprus	391	837	510	5.4	1.6	5.7	92.8	2.9
Czech Republic	1101	19446	460	11.6	1.3	11.7	103.6	7.3
Denmark	1388	13527	504	2.6	0.7	2.7	61.8	1.7
France	1071	151531	548	4.1	0.4	4.1	56.2	2.9
Germany	2296	262789	457	5.0	0.8	5.0	80.1	3.6
Greece	456	14620	539	6.7	2.4	7.2	112.7	7.3
Italy	398	104477	460	9.2	0.3	9.2	103.2	8.7
Lithuania	734	1360	547	2.3	1.7	2.8	84.6	2.7
Russian Federation	1638	42858	555	8.7	1.4	8.8	106.7	5.6
Slovenia	1536	22881	491	9.8	0.7	9.9	105.8	5.0
Sweden	1001	16408	523	4.7	0.7	4.7	88.4	3.2
Switzerland	1404	11343	514	5.1	1.1	5.2	87.5	2.6
United States	2785	496852	459	5.2	1.1	5.3	86.6	3.8

S.D. = standard deviation
S.E. = standard error

Table 5.7 Sampling and Imputation Standard Errors - Calculus Scale
Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	645	39498	530	11.6	1.1	11.7	100.3	11.2
Austria	782	31063	439	6.4	1.3	6.5	88.9	5.1
Canada	2781	58606	503	3.4	1.2	3.6	92.8	3.1
Cyprus	391	837	561	5.0	1.4	5.2	99.8	3.8
Czech Republic	1101	19446	446	9.7	0.7	9.7	97.3	7.5
Denmark	1388	13527	508	3.3	0.3	3.3	85.6	2.2
France	1071	151531	560	2.9	0.7	3.0	64.0	2.3
Germany	2296	262789	454	4.3	0.7	4.4	85.0	3.2
Greece	456	14620	538	6.8	2.8	7.3	98.3	6.4
Italy	398	104477	520	10.4	0.4	10.4	107.3	6.2
Lithuania	734	1360	498	2.0	1.5	2.5	78.1	2.2
Russian Federation	1638	42858	537	9.0	1.3	9.1	106.4	7.4
Slovenia	1536	22881	471	6.6	0.8	6.6	71.2	2.5
Sweden	1001	16408	480	4.3	0.7	4.4	88.0	2.7
Switzerland	1404	11343	512	5.6	1.1	5.7	96.6	3.5
United States	2785	496852	450	4.0	1.0	4.1	97.3	4.1

S.D. = standard deviation
S.E. = standard error

Table 5.8 Sampling and Imputation Standard Errors - Geometry Scale
Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	645	39498	496	12.4	1.9	12.5	122.5	10.0
Austria	782	31063	462	7.8	0.5	7.9	97.7	6.5
Canada	2781	58606	499	3.7	0.9	3.8	95.0	1.9
Cyprus	391	837	517	4.7	1.6	4.9	99.1	4.4
Czech Republic	1101	19446	494	9.8	0.9	9.8	102.2	8.7
Denmark	1388	13527	527	2.9	1.2	3.1	70.5	2.2
France	1071	151531	544	3.7	0.9	3.8	76.8	2.5
Germany	2296	262789	487	5.4	0.5	5.5	75.1	4.0
Greece	456	14620	498	8.0	3.3	8.7	116.3	7.2
Italy	398	104477	480	9.5	0.5	9.5	103.7	9.0
Lithuania	734	1360	515	2.3	1.6	2.8	82.3	1.9
Russian Federation	1638	42858	548	9.0	1.6	9.2	103.6	5.5
Slovenia	1536	22881	476	7.5	0.9	7.6	83.4	3.3
Sweden	1001	16408	492	4.4	0.6	4.4	83.0	2.6
Switzerland	1404	11343	547	4.1	1.1	4.2	87.0	3.1
United States	2785	496852	424	5.1	0.7	5.1	96.5	4.4

S.D. = standard deviation
S.E. = standard error

Table 5.9 Sampling and Imputation Standard Errors - Physics Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	661	31619	518	5.8	2.1	6.2	82.2	3.6
Austria	777	30795	435	6.4	0.8	6.4	83.3	4.6
Canada	2367	51179	485	2.8	1.8	3.3	86.7	3.0
Cyprus	368	837	494	5.7	0.8	5.8	105.3	5.3
Czech Republic	1087	19428	451	6.2	0.7	6.2	82.2	5.9
Denmark	654	2073	534	3.9	1.3	4.2	84.6	3.9
France	1110	151531	466	3.5	1.5	3.8	65.9	3.1
Germany	723	87888	522	11.8	1.2	11.9	94.2	5.3
Greece	459	14668	486	5.4	1.2	5.6	87.3	3.7
Latvia (LSS)	708	979	488	21.5	0.6	21.5	100.3	10.6
Norway	1048	4369	581	6.1	2.3	6.5	90.5	2.5
Russian Federation	1233	32975	545	11.4	2.4	11.6	110.3	5.0
Slovenia	747	11706	523	15.3	2.5	15.5	108.9	8.7
Sweden	1012	16459	573	3.8	0.8	3.9	92.1	2.8
Switzerland	1371	11276	488	3.4	0.8	3.5	88.3	2.9
United States	3114	522784	423	3.2	0.7	3.3	59.9	3.2

S.D. = standard deviation
S.E. = standard error

Table 5.10 Sampling and Imputation Standard Errors - Mechanics Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	661	31619	507	5.9	1.2	6.1	87.6	3.9
Austria	777	30795	420	4.9	0.2	4.9	78.0	3.5
Canada	2367	51179	473	3.5	0.8	3.6	89.0	2.8
Cyprus	368	837	530	6.6	0.4	6.6	116.9	5.4
Czech Republic	1087	19428	469	5.9	0.8	6.0	80.5	6.3
Denmark	654	2073	529	4.7	1.2	4.9	87.3	3.9
France	1110	151531	457	3.9	1.8	4.3	74.8	3.8
Germany	723	87888	495	9.2	1.7	9.4	90.3	6.6
Greece	459	14668	514	6.3	1.8	6.5	90.7	4.4
Latvia (LSS)	708	979	489	18.1	0.6	18.1	91.6	9.1
Norway	1048	4369	572	5.9	2.3	6.4	89.5	4.0
Russian Federation	1233	32975	537	9.3	0.6	9.3	91.4	6.5
Slovenia	747	11706	552	17.2	1.9	17.3	120.2	11.3
Sweden	1012	16459	563	4.0	0.4	4.0	80.4	2.6
Switzerland	1371	11276	482	3.5	0.3	3.5	86.2	3.5
United States	3114	522784	420	2.7	0.5	2.8	59.0	2.7

S.D. = standard deviation
S.E. = standard error

Table 5.11 Sampling and Imputation Standard Errors - Electricity and Magnetism Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	661	31619	512	4.2	1.1	4.4	91.0	5.5
Austria	777	30795	432	6.2	1.2	6.3	93.6	4.8
Canada	2367	51179	485	3.4	1.4	3.7	83.2	4.2
Cyprus	368	837	502	6.2	1.0	6.3	114.5	7.0
Czech Republic	1087	19428	465	5.4	0.7	5.5	75.2	6.0
Denmark	654	2073	513	3.5	1.5	3.8	79.4	4.6
France	1110	151531	494	4.0	0.4	4.1	59.9	3.3
Germany	723	87888	512	9.6	2.3	9.9	92.0	5.4
Greece	459	14668	520	6.4	1.7	6.6	105.1	4.8
Latvia (LSS)	708	979	485	17.3	1.1	17.4	94.3	8.2
Norway	1048	4369	565	5.9	1.8	6.2	93.0	3.3
Russian Federation	1233	32975	549	9.0	1.7	9.2	107.2	4.9
Slovenia	747	11706	509	14.6	1.5	14.6	109.9	11.0
Sweden	1012	16459	570	3.2	0.7	3.3	88.1	3.3
Switzerland	1371	11276	480	4.5	0.8	4.5	93.7	3.3
United States	3114	522784	420	2.9	0.7	3.0	58.6	2.2

S.D. = standard deviation
S.E. = standard error

Table 5.12 Sampling and Imputation Standard Errors - Heat Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	661	31619	517	4.1	1.2	4.3	82.5	3.9
Austria	777	30795	445	5.5	0.9	5.6	93.0	4.9
Canada	2367	51179	508	3.7	2.1	4.2	85.3	5.0
Cyprus	368	837	476	6.7	0.4	6.7	145.6	5.8
Czech Republic	1087	19428	488	4.6	0.8	4.7	80.1	6.0
Denmark	654	2073	512	4.0	1.6	4.3	98.2	6.4
France	1110	151531	491	3.2	1.0	3.4	68.7	3.1
Germany	723	87888	496	5.8	2.6	6.4	97.7	5.0
Greece	459	14668	481	7.1	1.4	7.2	117.7	6.4
Latvia (LSS)	708	979	504	21.3	1.3	21.4	112.8	9.3
Norway	1048	4369	536	3.7	2.1	4.3	70.2	2.6
Russian Federation	1233	32975	530	10.3	1.6	10.4	105.4	7.3
Slovenia	747	11706	521	10.2	1.9	10.4	117.8	8.3
Sweden	1012	16459	522	4.2	0.9	4.3	81.9	2.4
Switzerland	1371	11276	509	3.4	0.9	3.6	89.5	2.7
United States	3114	522784	477	2.9	0.7	3.0	58.4	2.8

S.D. = standard deviation
S.E. = standard error

Table 5.13 Sampling and Imputation Standard Errors - Wave Phenomena Scale Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	661	31619	519	6.6	1.8	6.9	98.0	6.0
Austria	777	30795	468	7.2	0.5	7.3	86.8	7.4
Canada	2367	51179	488	2.8	1.5	3.2	80.0	2.6
Cyprus	368	837	507	6.5	0.4	6.5	119.1	7.4
Czech Republic	1087	19428	447	5.4	0.4	5.4	75.9	3.4
Denmark	654	2073	537	5.2	1.7	5.5	97.1	5.6
France	1110	151531	463	3.6	0.2	3.6	73.2	2.7
Germany	723	87888	530	10.0	2.3	10.3	97.7	5.2
Greece	459	14668	453	5.2	0.9	5.3	93.3	4.9
Latvia (LSS)	708	979	498	17.5	1.0	17.6	91.1	11.9
Norway	1048	4369	560	5.1	1.7	5.4	90.4	2.9
Russian Federation	1233	32975	515	9.2	2.0	9.4	105.9	5.8
Slovenia	747	11706	514	11.4	1.4	11.5	115.1	7.0
Sweden	1012	16459	560	4.6	0.8	4.7	107.5	3.5
Switzerland	1371	11276	498	3.0	0.7	3.1	89.1	3.0
United States	3114	522784	451	2.1	0.6	2.2	53.3	1.6

S.D. = standard deviation
S.E. = standard error

Table 5.14 Sampling and Imputation Standard Errors Particle, Quantum, Astrophysics and Relativity Students in their Final Year of Secondary School

Country	Sample Size	Population Size	Mean of 5 Plausible Values	Error Due to Sampling	Error Due to Imputation	Sampling and Imputation Error	S.D.	S.E. of the S.D.
Australia	661	31619	521	5.6	1.3	5.8	88.0	4.0
Austria	777	30795	480	5.9	1.2	6.0	82.5	3.8
Canada	2367	51179	494	2.7	0.5	2.7	80.2	2.6
Cyprus	368	837	434	5.2	0.4	5.2	131.3	4.7
Czech Republic	1087	19428	453	4.9	0.9	4.9	87.3	4.1
Denmark	654	2073	544	4.7	1.2	4.9	81.4	4.7
France	1110	151531	474	3.3	0.7	3.4	60.9	4.0
Germany	723	87888	545	12.8	2.5	13.1	107.6	7.4
Greece	459	14668	447	4.9	0.6	4.9	92.8	5.3
Latvia (LSS)	708	979	488	19.0	1.3	19.0	95.3	9.8
Norway	1048	4369	576	5.1	1.4	5.3	84.0	4.2
Russian Federation	1233	32975	542	9.8	1.4	9.9	98.0	6.5
Slovenia	747	11706	511	15.0	1.9	15.1	112.8	10.1
Sweden	1012	16459	560	3.4	0.8	3.5	77.1	2.7
Switzerland	1371	11276	488	3.7	0.8	3.8	83.3	3.7
United States	3114	522784	456	2.4	0.6	2.5	49.3	2.9

S.D. = standard deviation
S.E. = standard error

REFERENCES

Adams, R. J., and Gonzalez, E. J. (1996). The TIMSS test design. In M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study technical report, volume I: Design and development*. Chestnut Hill, MA: Boston College.

Foy, P., Rust, K., and Schleicher, A. (1996). Sample design. In M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study technical report, volume I: Design and development*. Chestnut Hill, MA: Boston College.

Gonzalez, E. J., Smith, T. A., and Sibberns, H. (Eds.) (1998). *User guide for the TIMSS international database: Final year of secondary school – 1995 assessment.* Chestnut Hill, MA: Boston College.

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

Kish, L. (1965). *Survey sampling.* New York: John Wiley & Sons.

Mullis, I.V.S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., and Smith, T.A. (1998). *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study*. 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.