CHAPTER 4

# Estimating Standard Errors in the TIMSS Advanced 2015 Results

Pierre Foy
Sylvie LaRoche

To obtain estimates of students' proficiency in advanced mathematics and physics that are both accurate and cost-effective, TIMSS Advanced 2015 made extensive use of probability sampling techniques to sample students from student populations in their final year of secondary schooling, 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 additional 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 TIMSS Advanced 2015 International Results in Advanced Mathematics and Physics report 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 student populations, 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

TIMSS Advanced 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 TIMSS Advanced is a more challenging endeavor.

A common way to estimate the sampling variance in multi-stage cluster sampling designs is through resampling schemes such as the balanced repeated replication and Jackknife techniques

(Johnson & Rust, 1992; Wolter, 1985). TIMSS Advanced 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, totals, and percentages.

At the core of the JRR technique is the grouping of sampling units into zones based on sample design conditions (e.g., strata) and subsequent repeated draws of subsamples from these zones, i.e., repeated replication. For TIMSS Advanced, the two main features of the TIMSS Advanced 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 in each zone and by 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 for the purposes of calculating jackknife standard errors.[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 TIMSS Advanced 2015 participating country.

**Exhibit 4.1: Number of Sampling Zones for Each TIMSS Advanced 2015 Participating Country**

| Country | TIMSS Advanced 2015 Sampling Zones | |
|---|---|---|
| | Advanced Mathematics | Physics |
| France | 73 | 73 |
| Italy | 58 | 58 |
| Lebanon | 75 | 75 |
| Norway | 75 | 64 |
| Portugal | 75 | 75 |
| Russian Federation | 73 | 50 |
| Russian Federation 6hr+[3] | 73 | – |
| Slovenia | 75 | 58 |
| Sweden | 71 | 69 |
| United States | 75 | 75 |

A dash (-) indicates comparable data are not available.

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 If a remaining school consists of 2 sampled classrooms, each classroom becomes a "quasi" school.

3 For advanced mathematics, the Russian Federation participated in 2015 with an expanded population that included the more specialized students assessed in 1995 and 2008.

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. In both subsamples, all students in the other sampling zones are included. When a school is removed from the sample, the weights of the remaining school are doubled to make up for the omitted school. 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.[34]

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

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

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

$$W_{hji} = k_{hj} \cdot W_{0i} \quad (2)$$

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.

4 Prior to 2015, TIMSS Advanced used 75 subsamples and sets of replicate weights to calculate the JRR sampling variances. To provide more accurate estimates, starting in 2015 TIMSS Advanced 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 |

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.

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 TIMSS Advanced JRR algorithm, is given by the following equation:

$$Var_{jrr}(t_0) = \frac{1}{2} \sum_{h=1}^{75} \sum_{j=1}^{2} (t_{hj} - t_0)^2 \qquad (3)$$

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 $h$ ($h = 1, \ldots ,75$), where the $j^{th}$ school (1st or 2nd) in the zone is included and the other removed.

The sampling variance estimated with the TIMSS Advanced JRR method properly measures 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 TIMSS Advanced *Students Like Learning Advanced Mathematics* contextual scale, and the percentage of students with at least one parent with a university degree.

## 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 4 of the [TIMSS Advanced 2015 Assessment Frameworks](), the TIMSS Advanced 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 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, TIMSS Advanced 2015 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:

$$t_0 = \frac{1}{5}\sum_{m=1}^{5} t_m \tag{4}$$

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

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

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:

$$Var_{tot}(t_0) = Var_{jrr}(t_0) + Var_{imp}(t_0) \tag{6}$$

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

$$Var_{jrr}(t_0) = \frac{1}{5}\sum_{m=1}^{5} Var_{jrr}(t_m) \tag{7}$$

where

$$Var_{jrr}(t_m) = \frac{1}{2}\sum_{h=1}^{75}\sum_{j=1}^{2}(t_{mhj} - t_m)^2 \tag{8}$$

and $t_{mhj}$ is the appropriate JRR estimate based on plausible value $m$ 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 TIMSS Advanced mathematics achievement for girls and the percentage of students who reach the TIMSS Advanced intermediate international benchmark of physics achievement.

Appendices 4A and 4B provide details on the jackknife sampling variance, the imputation variance, the total variance, and the standard error for each country's mean proficiency estimates in advanced mathematics and physics, respectively.

## Estimating Standard Errors for International Averages

Some exhibits in the TIMSS Advanced 2015 reports include international averages and their standard errors. For example, Exhibit M4.1 reports the international average for the percentages of students in three categories of home educational resources and their advanced mathematics achievement. International averages are computed using the data from the participating countries included in the main table of an exhibit. Data from the benchmarking participants is not included in the estimation of international averages.

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

$$t_{int} = \frac{1}{N}\sum_{i=1}^{N} t_{0i} \tag{9}$$

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 average $t_{int}$ is given by:

$$Var(t_{int}) = \frac{1}{N^2} \sum_{i=1}^{N} Var_{tot}(t_{0i}) \tag{10}$$

where $Var_{tot}(t_{0i})$ is the total variance of our statistic of interest for the $i^{th}$ country, as given in equation (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 (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 Advanced Mathematics

### Summary Statistics and Standard Errors for Proficiency in Overall Advanced Mathematics

| Country | Sample Size | Overall Advanced Mathematics | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3967 | 462.664 | 7.014 | 2.338 | 9.352 | 3.058 |
| Italy | 3318 | 421.944 | 24.051 | 4.466 | 28.517 | 5.340 |
| Lebanon | 1161 | 531.874 | 7.817 | 1.575 | 9.392 | 3.065 |
| Norway | 2537 | 459.209 | 18.505 | 2.510 | 21.015 | 4.584 |
| Portugal | 4068 | 482.253 | 5.552 | 0.687 | 6.239 | 2.498 |
| Russian Federation | 7558 | 484.662 | 31.928 | 0.773 | 32.701 | 5.718 |
| Russian Federation 6hr+[1] | 3431 | 540.095 | 59.743 | 1.604 | 61.347 | 7.832 |
| Slovenia | 2922 | 459.794 | 10.915 | 0.635 | 11.550 | 3.398 |
| Sweden | 3937 | 431.082 | 13.203 | 3.057 | 16.260 | 4.032 |
| United States | 2954 | 484.984 | 25.501 | 1.261 | 26.762 | 5.173 |

### Summary Statistics and Standard Errors for Proficiency in Algebra

| Country | Sample Size | Algebra | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3967 | 469.180 | 7.722 | 0.935 | 8.657 | 2.942 |
| Italy | 3318 | 414.401 | 24.501 | 1.329 | 25.830 | 5.082 |
| Lebanon | 1161 | 525.424 | 9.492 | 6.391 | 15.882 | 3.985 |
| Norway | 2537 | 446.266 | 16.147 | 0.639 | 16.787 | 4.097 |
| Portugal | 4068 | 494.572 | 5.655 | 1.900 | 7.555 | 2.749 |
| Russian Federation | 7558 | 494.682 | 38.602 | 0.533 | 39.135 | 6.256 |
| Russian Federation 6hr+ | 3431 | 556.293 | 72.822 | 7.897 | 80.719 | 8.984 |
| Slovenia | 2922 | 473.548 | 10.763 | 1.331 | 12.095 | 3.478 |
| Sweden | 3937 | 422.083 | 13.076 | 3.442 | 16.518 | 4.064 |
| United States | 2954 | 478.150 | 23.399 | 1.929 | 25.327 | 5.033 |

1 For advanced mathematics, the Russian Federation participated in 2015 with an expanded population that included the more specialized students assessed in 1995 and 2008.

### Summary Statistics and Standard Errors for Proficiency in Calculus

| Country | Sample Size | Calculus | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3967 | 465.938 | 7.852 | 2.286 | 10.138 | 3.184 |
| Italy | 3318 | 432.688 | 23.062 | 4.488 | 27.550 | 5.249 |
| Lebanon | 1161 | 543.753 | 7.425 | 7.841 | 15.266 | 3.907 |
| Norway | 2537 | 463.304 | 23.197 | 4.898 | 28.095 | 5.300 |
| Portugal | 4068 | 475.762 | 6.307 | 0.657 | 6.964 | 2.639 |
| Russian Federation | 7558 | 459.057 | 33.355 | 1.103 | 34.459 | 5.870 |
| Russian Federation 6hr+ | 3431 | 513.019 | 59.430 | 4.189 | 63.619 | 7.976 |
| Slovenia | 2922 | 436.630 | 13.657 | 5.376 | 19.033 | 4.363 |
| Sweden | 3937 | 438.151 | 14.000 | 0.877 | 14.877 | 3.857 |
| United States | 2954 | 504.215 | 31.116 | 4.456 | 35.572 | 5.964 |

### Summary Statistics and Standard Errors for Proficiency in Geometry

| Country | Sample Size | Geometry | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3967 | 440.750 | 6.582 | 6.742 | 13.324 | 3.650 |
| Italy | 3318 | 413.154 | 25.306 | 7.307 | 32.614 | 5.711 |
| Lebanon | 1161 | 525.577 | 11.947 | 1.583 | 13.530 | 3.678 |
| Norway | 2537 | 472.777 | 16.784 | 4.583 | 21.367 | 4.622 |
| Portugal | 4068 | 463.763 | 7.048 | 3.213 | 10.261 | 3.203 |
| Russian Federation | 7558 | 499.940 | 32.884 | 1.136 | 34.020 | 5.833 |
| Russian Federation 6hr+ | 3431 | 559.895 | 61.753 | 7.986 | 69.740 | 8.351 |
| Slovenia | 2922 | 455.854 | 14.264 | 1.495 | 15.759 | 3.970 |
| Sweden | 3937 | 429.982 | 11.180 | 2.227 | 13.407 | 3.662 |
| United States | 2954 | 454.953 | 27.669 | 4.330 | 32.000 | 5.657 |

### Summary Statistics and Standard Errors for Proficiency in Advanced Mathematics Knowing

| Country | Sample Size | Advanced Mathematics Knowing | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3967 | 475.334 | 6.727 | 0.521 | 7.248 | 2.692 |
| Italy | 3318 | 422.828 | 23.893 | 6.503 | 30.396 | 5.513 |
| Lebanon | 1161 | 542.963 | 7.240 | 12.813 | 20.054 | 4.478 |
| Norway | 2537 | 445.282 | 15.335 | 1.593 | 16.927 | 4.114 |
| Portugal | 4068 | 479.314 | 4.978 | 3.929 | 8.907 | 2.985 |
| Russian Federation | 7558 | 477.578 | 41.694 | 2.790 | 44.484 | 6.670 |
| Russian Federation 6hr+ | 3431 | 537.612 | 77.292 | 0.444 | 77.736 | 8.817 |
| Slovenia | 2922 | 466.074 | 12.012 | 0.503 | 12.514 | 3.538 |
| Sweden | 3937 | 404.981 | 16.326 | 5.474 | 21.800 | 4.669 |
| United States | 2954 | 487.789 | 30.660 | 1.869 | 32.529 | 5.703 |

### Summary Statistics and Standard Errors for Advanced Mathematics Applying

| Country | Sample Size | Advanced Mathematics Applying | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3967 | 448.880 | 8.200 | 3.488 | 11.688 | 3.419 |
| Italy | 3318 | 425.189 | 23.585 | 5.751 | 29.336 | 5.416 |
| Lebanon | 1161 | 529.261 | 8.395 | 6.287 | 14.682 | 3.832 |
| Norway | 2537 | 458.975 | 19.848 | 6.366 | 26.214 | 5.120 |
| Portugal | 4068 | 475.842 | 5.482 | 2.863 | 8.345 | 2.889 |
| Russian Federation | 7558 | 490.747 | 35.297 | 2.024 | 37.321 | 6.109 |
| Russian Federation 6hr+ | 3431 | 543.948 | 61.129 | 4.556 | 65.685 | 8.105 |
| Slovenia | 2922 | 464.915 | 10.022 | 5.620 | 15.642 | 3.955 |
| Sweden | 3937 | 434.041 | 11.844 | 1.138 | 12.981 | 3.603 |
| United States | 2954 | 479.594 | 28.578 | 1.551 | 30.129 | 5.489 |

### Summary Statistics and Standard Errors for Proficiency in Advanced Mathematics Reasoning

| Country | Sample Size | Advanced Mathematics Reasoning | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3967 | 462.245 | 7.175 | 2.539 | 9.715 | 3.117 |
| Italy | 3318 | 411.144 | 26.722 | 8.630 | 35.352 | 5.946 |
| Lebanon | 1161 | 526.658 | 9.979 | 5.610 | 15.589 | 3.948 |
| Norway | 2537 | 468.619 | 17.332 | 2.028 | 19.361 | 4.400 |
| Portugal | 4068 | 487.964 | 6.896 | 5.330 | 12.226 | 3.497 |
| Russian Federation | 7558 | 484.022 | 26.849 | 0.964 | 27.813 | 5.274 |
| Russian Federation 6hr+ | 3431 | 540.789 | 46.446 | 5.724 | 52.170 | 7.223 |
| Slovenia | 2922 | 442.450 | 13.550 | 2.173 | 15.723 | 3.965 |
| Sweden | 3937 | 446.651 | 11.392 | 3.524 | 14.916 | 3.862 |
| United States | 2954 | 484.454 | 24.881 | 3.218 | 28.099 | 5.301 |

# Appendix 4B: Summary Statistics and Standard Errors for Proficiency in Physics

### Summary Statistics and Standard Errors for Proficiency in Overall Physics

| Country | Sample Size | Overall Physics | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3958 | 373.057 | 13.761 | 2.110 | 15.871 | 3.984 |
| Italy | 3424 | 373.925 | 43.465 | 4.287 | 47.753 | 6.910 |
| Lebanon | 1156 | 410.159 | 13.009 | 7.231 | 20.240 | 4.499 |
| Norway | 2472 | 507.262 | 19.351 | 1.523 | 20.874 | 4.569 |
| Portugal | 1783 | 466.609 | 19.187 | 2.186 | 21.373 | 4.623 |
| Russian Federation | 3822 | 507.534 | 48.865 | 0.923 | 49.788 | 7.056 |
| Slovenia | 1106 | 531.033 | 5.624 | 0.741 | 6.365 | 2.523 |
| Sweden | 3727 | 454.667 | 33.430 | 1.327 | 34.756 | 5.895 |
| United States | 2932 | 437.338 | 91.101 | 2.411 | 93.512 | 9.670 |

### Summary Statistics and Standard Errors for Proficiency in Mechanics and Thermodynamics

| Country | Sample Size | Mechanics and Thermodynamics | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3958 | 327.429 | 21.783 | 10.630 | 32.413 | 5.693 |
| Italy | 3424 | 375.967 | 40.310 | 0.553 | 40.863 | 6.392 |
| Lebanon | 1156 | 395.429 | 12.283 | 7.401 | 19.685 | 4.437 |
| Norway | 2472 | 502.581 | 14.897 | 1.834 | 16.731 | 4.090 |
| Portugal | 1783 | 489.030 | 19.817 | 3.639 | 23.456 | 4.843 |
| Russian Federation | 3822 | 514.150 | 44.622 | 0.524 | 45.146 | 6.719 |
| Slovenia | 1106 | 541.431 | 5.662 | 1.363 | 7.025 | 2.651 |
| Sweden | 3727 | 455.121 | 35.172 | 2.015 | 37.188 | 6.098 |
| United States | 2932 | 462.238 | 88.973 | 3.677 | 92.650 | 9.625 |

### Summary Statistics and Standard Errors for Proficiency in Electricity and Magnetism

| Country | Sample Size | Electricity and Magnetism | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3958 | 339.441 | 12.889 | 8.845 | 21.734 | 4.662 |
| Italy | 3424 | 425.180 | 40.728 | 2.913 | 43.642 | 6.606 |
| Lebanon | 1156 | 399.127 | 16.801 | 9.928 | 26.729 | 5.170 |
| Norway | 2472 | 514.368 | 25.508 | 4.344 | 29.852 | 5.464 |
| Portugal | 1783 | 431.308 | 21.707 | 12.508 | 34.215 | 5.849 |
| Russian Federation | 3822 | 515.395 | 58.447 | 6.236 | 64.684 | 8.043 |
| Slovenia | 1106 | 530.263 | 7.431 | 10.767 | 18.198 | 4.266 |
| Sweden | 3727 | 455.336 | 34.133 | 1.319 | 35.452 | 5.954 |
| United States | 2932 | 379.529 | 144.254 | 4.934 | 149.187 | 12.214 |

### Summary Statistics and Standard Errors for Proficiency in Wave Phenomena and Atomic/Nuclear Physics

| Country | Sample Size | Wave Phenomena and Atomic/Nuclear Physics | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3958 | 417.815 | 14.584 | 5.345 | 19.928 | 4.464 |
| Italy | 3424 | 329.071 | 54.069 | 8.577 | 62.646 | 7.915 |
| Lebanon | 1156 | 430.516 | 16.397 | 30.229 | 46.626 | 6.828 |
| Norway | 2472 | 507.447 | 25.546 | 1.899 | 27.445 | 5.239 |
| Portugal | 1783 | 455.513 | 20.954 | 16.939 | 37.893 | 6.156 |
| Russian Federation | 3822 | 490.105 | 54.418 | 1.300 | 55.718 | 7.464 |
| Slovenia | 1106 | 510.914 | 8.774 | 11.558 | 20.332 | 4.509 |
| Sweden | 3727 | 450.915 | 38.095 | 1.638 | 39.733 | 6.303 |
| United States | 2932 | 430.688 | 75.283 | 0.952 | 76.235 | 8.731 |

### Summary Statistics and Standard Errors for Proficiency in Physics Knowing

| Country | Sample Size | Physics Knowing | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3958 | 374.787 | 14.150 | 0.928 | 15.078 | 3.883 |
| Italy | 3424 | 367.266 | 38.479 | 5.396 | 43.874 | 6.624 |
| Lebanon | 1156 | 378.083 | 17.288 | 4.842 | 22.130 | 4.704 |
| Norway | 2472 | 529.308 | 14.771 | 2.728 | 17.499 | 4.183 |
| Portugal | 1783 | 474.028 | 19.368 | 2.277 | 21.646 | 4.653 |
| Russian Federation | 3822 | 516.718 | 50.279 | 6.026 | 56.305 | 7.504 |
| Slovenia | 1106 | 521.009 | 8.776 | 8.695 | 17.472 | 4.180 |
| Sweden | 3727 | 451.645 | 33.491 | 2.964 | 36.455 | 6.038 |
| United States | 2932 | 444.055 | 88.814 | 6.639 | 95.453 | 9.770 |

### Summary Statistics and Standard Errors for Proficiency in Physics Applying

| Country | Sample Size | Physics Applying | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3958 | 358.066 | 14.742 | 16.803 | 31.545 | 5.617 |
| Italy | 3424 | 371.133 | 47.889 | 6.133 | 54.022 | 7.350 |
| Lebanon | 1156 | 432.517 | 14.148 | 15.471 | 29.618 | 5.442 |
| Norway | 2472 | 484.133 | 24.229 | 3.701 | 27.930 | 5.285 |
| Portugal | 1783 | 452.100 | 20.317 | 11.804 | 32.121 | 5.668 |
| Russian Federation | 3822 | 508.196 | 55.652 | 1.411 | 57.063 | 7.554 |
| Slovenia | 1106 | 543.490 | 6.340 | 8.035 | 14.374 | 3.791 |
| Sweden | 3727 | 454.326 | 36.800 | 3.648 | 40.448 | 6.360 |
| United States | 2932 | 420.403 | 100.159 | 3.241 | 103.400 | 10.169 |

### Summary Statistics and Standard Errors for Proficiency in Physics Reasoning

| Country | Sample Size | Physics Reasoning | | | | |
|---|---|---|---|---|---|---|
| | | Mean Proficiency | Jackknife Sampling Variance | Imputation Variance | Total Variance | Overall Standard Error |
| France | 3958 | 396.765 | 15.028 | 2.986 | 18.013 | 4.244 |
| Italy | 3424 | 374.825 | 50.115 | 2.628 | 52.743 | 7.262 |
| Lebanon | 1156 | 374.762 | 19.943 | 18.998 | 38.941 | 6.240 |
| Norway | 2472 | 518.935 | 19.645 | 12.832 | 32.477 | 5.699 |
| Portugal | 1783 | 480.645 | 14.479 | 0.525 | 15.004 | 3.873 |
| Russian Federation | 3822 | 493.011 | 42.047 | 2.365 | 44.412 | 6.664 |
| Slovenia | 1106 | 514.024 | 9.735 | 23.082 | 32.818 | 5.729 |
| Sweden | 3727 | 450.388 | 34.302 | 3.564 | 37.866 | 6.154 |
| United States | 2932 | 454.761 | 76.075 | 1.210 | 77.285 | 8.791 |