4 Calculation of Sampling Weights

Jean Dumais
Pierre Foy
Statistics Canada

4.1 INTRODUCTION

This chapter describes the procedures for computing sampling weights for the TIMSS assessment of students in the final year of secondary school (Population 3), in which 24 countries took part. TIMSS Population 3 is defined as all students in their final year of secondary education, that is, all students who upon successful completion of that final school year would either enter the labor market or tertiary education. This definition is meant to be as inclusive as possible.

The TIMSS sampling design was intended to provide estimates of the mathematics and science literacy of all students in the final year of secondary school, while also assessing the advanced mathematics and physics knowledge of students with preparation in these subjects. In addition to characterizing the entire population of final-year students, therefore, the design had to produce accurate estimates of two overlapping subpopulations: students with preparation in advanced mathematics, and students with preparation in physics. In several countries where the overlap was complete (all students that study advanced mathematics also study physics) there were just two groups, those that studied advanced mathematics and physics and those that did not. In countries with clearly defined tracks for upper secondary students, these two groups were often in different schools, which further simplified the sampling procedure. However, in general the situation was more complicated, and a more complex design was required. This design is summarized below, and is described in more detail in Chapter 2.

An essential aspect of the sampling design was that students were stratified according to their level of preparation in mathematics and physics, so that appropriate test booklets could be assigned to them. As described in Chapter 2, each student was characterized as having taken advanced mathematics (M) or not (O), and as having taken physics (P) or not (O). Combining these two-way classifications yields four mutually exclusive and exhaustive categories of students:

OO Students having studied neither advanced mathematics nor physics

OP Students having studied physics but not advanced mathematics

MO Students having studied advanced mathematics but not physics

MP Students having studied both advanced mathematics and physics

In some school systems, students in each group were readily identifiable by virtue of their track assignment or school type, whereas in others it was necessary to categorize individual students in terms of their course-taking history.

Four kinds of student test booklets were assigned to students on the basis of this classification, so that each student completed one 90-minute test booklet. Students classified as OO received either booklet 1A or 1B, the two booklets containing items related to mathematics and science literacy. Students classified as OP received either booklet 1A or 1B, or one of the three booklets containing physics material (2A, 2B, or 2C). Students classified as MO received either booklet 1A or 1B, or one of the three booklets containing advanced mathematics material (3A, 3B, or 3C). Students classified as MP also received one booklet, which could be 1A, 1B, 2A, 2B, 2C, 3A, 3B, 3C or 4. Booklet 4 contained mathematics and science literacy, advanced mathematics, and physics items.

In reporting the achievement of advanced mathematics students, the sample consisted of all MO and MP students that were assigned one of the mathematics booklets (booklets 3A, 3B, or 3C) or the combined literacy, advanced mathematics, and physics booklet (booklet 4). Similarly, the sample of physics students consisted of all OP and MP students that were assigned one of the physics booklets (booklets 2A, 2B, or 2C) or the combined booklet (booklet 4). The sample for reporting on mathematics and science literacy comprised all students in each of the OO, OP, MO, and MP strata that were assigned one of the literacy booklets (booklets 1A or 1B) or the combined booklet (booklet 4).

The basic sample design (intended for use in comprehensive systems where all four kinds of students could be found in all schools) was straightforward. It consisted of a two-stage procedure where schools were sampled with probability proportional to size in the first stage, and an equal number of OO, OP, MO, and MP students was sampled in the second stage. However, implementation varied from country to country, depending on the structure of the education system, and was often quite complex. Some chose to stratify the schools explicitly, others did not; some sampled individual students while others preferred to sample entire classrooms. Details on sampling plans for individual countries are provided in Appendix B of this report. Given the number of variations on the basic design, and the frequent necessity of using different sampling fractions for each student type, the derivation of appropriate sampling weights was a very important step in ensuring the computation of proper survey estimates.

TIMSS made use of item response theory (IRT) methods to derive scales for mathematics and science literacy, advanced mathematics, and physics. The IRT methodology provides an estimate of the proficiency on the scale for each student, even though each student completed only one booklet and hence responded to only part of the assessment item pool. For example, the literacy scale is based on the contents of booklets 1A and 1B, but students in the literacy sample completed just one of these booklets. Most of the international reporting is by scale, so it was necessary to have sampling weights appropriate to this level, i.e., for students that took either booklet 1A or 1B. However, TIMSS also reports student performance on selected individual items, and these are specific to particular booklets. Consequently, it was necessary to compute sampling weights that could be used at the booklet level also, i.e., for the students that completed booklet 1A, or for the students that completed booklet 1B.

4.2 GENERAL WEIGHTING PROCEDURE

Although the basic sampling design specified just a two-stage procedure, since participants could sample either intact classes or individual students it was convenient for computational reasons to think of classes and students as separate sampling stages; and the distribution of booklets at random within schools effectively adds another sampling stage. Computationally, therefore, the sampling weights were assembled in four steps which reflected the multi-stage nature of the sample design. The first step produced a school-level weighting factor. The second step produced a classroom-level weighting factor. The third step produced a student-level weighting factor. The last step produced booklet-level and a scale-level weights. Non-response adjustments were also made to the weighting factors. The overall estimation weight attached to each student record was the product of the four intermediate weights: the school weight, the classroom weight, the student weight, and the scale or booklet weight.

4.2.1 The School Weight

The school weight represents the inverse of the first-stage selection probability of a sampled school. The TIMSS design requires that school selection probabilities be proportional to the school size (PPS), defined as enrollment in the target population. Participants were encouraged to stratify schools explicitly by factors that would improve the precision of the sample or guarantee coverage of special populations. This was in addition to the requirement to stratify students by academic preparation so as to identify the reporting populations. The basic school weight for the *i*th sampled school in a given explicit stratum is thus defined as

$$BW_i^{sc} = \frac{M}{n \times m_i}$$

where n is the number of sampled schools in the stratum, m_i is the measure of size for the *i*th school, and

$$M = \sum_{i=1}^{N} m_i$$

where N is the total number of schools in the stratum.

A few countries opted for simple random sampling of schools rather than PPS; this means that every school has the same unit size ($m_i \equiv 1$) and that M = N.

In two large participating countries (the United States and the Russian Federation) it was necessary to introduce an extra stage of sampling whereby geographical regions were sampled prior to sampling schools. For those countries the basic school weight incorporated a weighting factor to reflect this additional front-end sampling stage.

This weighting factor was calculated in the same way as the school weight since the geographical regions were also sampled with probability proportional to size. The resulting school weight was simply the product of the "region" weight and the school weight as described earlier.

The basic school weight was adjusted to reflect non-response among sampled schools. From the originally selected sample of n schools, occasionally schools were unable or unwilling to take part in the assessment. Whenever possible, these schools were replaced with replacement schools selected at the same time as the originals. In the end, the number of participating schools, n_p say, was sometimes smaller than the planned school sample size. Therefore the basic weight was adjusted to account for the reduction in sample size.

The school-level adjustment for non-response was calculated as follows within each explicit stratum:

$$A^{sc} = \frac{n}{n_p}$$

and the final school weight for the ith school thus becomes

$$FW_i^{sc} = BW_i^{sc} \times A^{sc}.$$

4.2.2 The Classroom Weight

The classroom weight is the inverse of the probability of selection of a sampled classroom within a sampled school. For many of the participants, the classroom weight was irrelevant since students were sampled directly within the school, in accordance with the basic sampling design, rather than via a sampled classroom. In such cases, the classroom weight was simply set at one (1.0). Classroom sampling was used only when all the students in the class belonged to the same sub-population, and consequently classroom weights were calculated independently for each sub-population.

For sub-population g within the ith school, let C_{gi} be the total number of classrooms. In most cases, one classroom only was selected with equal probability, and so the probability of selection was one divided by C_{gi}, and the reciprocal of this probability is the classroom weight. In those schools, the classroom weight assigned to the classroom from sub-population g in the ith school was

$$FW_{gi}^{cl} = C_{gi}$$

In a few instances, countries chose more than one classroom to better represent certain sub-populations. If c_{gi} is the number of classrooms selected at random, then

$$FW_{gi}^{cl} = \frac{C_{gi}}{c_{gi}}.$$

4.2.3 The Student Weight

The student weight is the inverse of the probability that a student within a sampled school or classroom will be sampled for the TIMSS testing. Let the number of enrolled students (after removing students that were out-of-scope or excluded) in school i and sub-population g (perhaps in a classroom) be N_{gi}. If the sample size is n_{gi}, then the basic student weight[1] is

$$BW_{gi}^{st} = \frac{N_{gi}}{n_{gi}}.$$

Occasionally a sampled student did not take part in the assessment, because of absence through illness or for some other reason, and so it was necessary to have a correction for student non-response. If there were r_{gi} students that responded, then the student non-response adjustment is

$$A_{gi}^{st} = \frac{n_{gi}}{r_{gi}}$$

and the final student-level estimation weight is:

$$FW_{gi}^{st} = BW_{gi}^{st} \times A_{gi}^{st}.$$

4.2.4 The Booklet Weights

Each sampled student was randomly assigned one of the nine test booklets. The possibilities for booklet assignment varied across sub-populations: OO students could receive one of the booklet 1 series only (the mathematics and science literacy booklets 1A or 1B); OP students could receive one of the booklet 2 series (the physics booklets 2A, 2B, or 2C) or one of the booklet 1 series (since all students are eligible to receive a literacy booklet); MO students could receive one of the booklet 3 series (the advanced mathematics booklets 3A, 3B, or 3C) or one of the booklet 1 series; and MP students could receive booklet 4 (which combines literacy, advanced mathematics, and physics questions) or any of the booklets in series 1, 2, or 3. The random assignment (or rotation, since booklets were actually distributed systematically within schools or classes) of booklets to students constituted another stage of sampling, and consequently had to be included in the calculation of weights. The booklet weights represent the booklet assignments as they were implemented in the student sample. There is one weight for each booklet series distributed within each sub-population. The set of students that were assigned the same booklet can be thought of as a sub-sample of the total sample. The booklet weights may be used when the focus of the analysis is on individual items rather than on summary scales.

[1] When classroom sampling was used, all students in the class were included and in that case $BW_{gi}^{st} = 1.0$.

To compute the booklet weights, we need to know for each sub-population g (g = OO, OP, MO, MP) how many booklets of each kind were distributed among the r_{gi} participants. Let r_{gi}^{b} be the number of participants in school i and sub-population g who received booklet b (b = 1, 2, 3, 4). Then we have:

$$W_{gi}^{1} = \frac{r_{gi}}{r_{gi}^{1}}\text{, for all } g$$

for the mathematics and science literacy booklets, 1A and 1B,

$$W_{gi}^{2} = \begin{cases} 0, \text{ if } g \in \{\text{OO, MO}\} \\ \dfrac{r_{gi}}{r_{gi}^{2}}, \text{ if } g \in \{\text{OP, MP}\} \end{cases}$$

for the physics booklets, 2A, 2B and 2C,

$$W_{gi}^{3} = \begin{cases} 0, \text{ if } g \in \{\text{OO, MO}\} \\ \dfrac{r_{gi}}{r_{gi}^{3}}, \text{ if } g \in \{\text{MO, MP}\} \end{cases}$$

for the advanced mathematics booklets, 3A, 3B, 3C, and finally

$$W_{gi}^{4} = \begin{cases} 0, \text{ if } g \in \{\text{OO, OP, MO}\} \\ \dfrac{r_{gi}}{r_{gi}^{4}}, \text{ if } g \in \{\text{MP}\} \end{cases}$$

for the combined booklet, booklet 4.

4.2.5 The Scale Weights

The booklet weights permit properly weighted analyses of student responses to individual items and were necessary since such analyses are an important aspect of the international reports. However, most of the TIMSS reporting made use of IRT scales, which summarize student performance across all of the items in a subject area. Because TIMSS requires more than one booklet to cover a subject area, but each student responded to only one booklet, the IRT scales had to combine item responses from different booklets, and hence from different students, and the scale weights had to reflect this. The scale weights are rooted in the booklet sub-samples and in the sub-populations.

The mathematics and science literacy estimation weight was based on all students that were assigned booklet series 1 or 4, and was constructed as follows:

$$W_{gi}^{MSL} = \frac{r_{gi}}{r_{gi}^{1,4}}\text{, for all } g$$

The weight for the physics scale involved students in the OP and MP sub-populations that were assigned booklet series 2 or 4, as follows:

$$W_{gi}^{P} = \begin{cases} 0, \text{ if } g \in \{\text{OO, MO}\} \\ \dfrac{r_{gi}}{r_{gi}^{2,4}}, \text{ if } g \in \{\text{OP, MP}\} \end{cases}$$

Finally, the weight for the advanced mathematics scale involved students in the MO and MP sub-populations that were assigned booklet series 3 or 4, as follows:

$$W_{gi}^{AM} = \begin{cases} 0, \text{ if } g \in \{\text{OO, OP}\} \\ \dfrac{r_{gi}}{r_{gi}^{3,4}}, \text{ if } g \in \{\text{MO, MP}\} \end{cases}$$

4.2.6 The Adjustments for Unbalanced Booklet Rotation

In many instances, there were fewer students from a sub-population in a school or class than the number of different booklets to be rotated. Uncorrected, this situation could make estimates of the population size vary with the choice of weight series. Since the estimated number of physics students (say) should be the same regardless of whether it was estimated using the “booklet weight” or the “scale weight,” this is not a desirable situation. Adjustment factors for booklet and scale weights were devised to correct for the potential relative rarity of certain booklets in the sample.

First, an estimate of the size of each sub-population g, g=OO, OP, MO, MP, was computed:

$$SI\hat{Z}E_g = \sum_{i}^{n}\sum_{j}^{r_{gl}} FW_i^{sc} \times FW_{gi}^{cl} \times FW_{gi}^{st}.$$

Then, an estimate was constructed of the size of the sub-population g using in turn each weight series b (b = 1, 2, 3, 4, MSL, P, AM) as defined in the preceding sections:

$$SI\hat{Z}E_g^{\mathrm{b}} = \sum_{i}^{n}\sum_{j}^{r_{gi}} FW_i^{sc} \times FW_{gi}^{cl} \times FW_{gi}^{st} \times W_{gi}^{b}.$$

The correction factor is therefore:

$$K_g^b = \frac{SI\hat{Z}E_g}{SI\hat{Z}E_g^{\mathrm{b}}}$$

for each booklet series b (b = 1, 2, 3, 4, MSL, P, AM) and each sub-population g, (g=OO, OP, MO, MP).

Hence, the final booklet or scale weight becomes:

$$FW_{gi}^{\mathrm{b}} = W_{gi}^{\mathrm{b}} \times K_g^{\mathrm{b}}.$$

4.2.7 The Complete Weight

At the end of the process, the estimation weight assigned to a student j depends on the school the student attends, the classroom, if classroom sampling has been used, the sub-population the student belongs to, and the booklet the student was assigned. Both booklet-based and scale-based weights were computed.

$$CW_{gij}^{\mathrm{b}} = FW_i^{sc} \times FW_{gi}^{cl} \times FW_{gi}^{st} \times FW_{gi}^{\mathrm{b}}.$$

Further details of the weights that were computed and are available in the TIMSS user database may be found in Gonzalez, Smith, and Sibberns (1998).

REFERENCES

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.