

TIMSS and PIRLS 2011

School Sampling Design Options to Accommodate Other Samples

Marc Joncas

Because TIMSS and PIRLS both were conducted in 2011, participating countries could choose to assess a variety of population combinations, including TIMSS at fourth and eighth grades, and PIRLS or prePIRLS at fourth grade. Therefore, participation could require a country to design not just one, but two or even three samples of schools and students. Depending on the circumstances, countries might wish to maximize the extent to which the same schools were included in all samples, or alternatively, ensure that no school appeared in more than one sample. In addition, countries participating in other studies at the same time as in TIMSS and/or PIRLS (national assessments, for example) needed to ensure that the TIMSS or PIRLS sampling design allowed them to avoid schools sampled for these studies.

To provide options for countries in designing their school samples, Statistics Canada implemented two special sampling procedures. Method A was applied when data collection occurred simultaneously for two or more populations (as was the case in 2011 with TIMSS 4th grade, TIMSS 8th grade and PIRLS), and the country wanted to control the overlap between the schools. Method B was used primarily to ensure that the TIMSS and/or PIRLS samples avoided schools sampled for other studies.

Sampling Method A: Sampling Modifications for Simultaneous Data Collection

This procedure stratifies the school population according to whether schools contain students from both populations to be sampled (fourth and eighth grades, for example), or students from one population only (fourth grade only or eighth grade only) as a way of controlling sample overlap. Each school is assigned a measure of size (MOS) based on the number of students in the two populations combined. Schools are sampled according to the sampling design described in the [Sample Design in TIMSS and PIRLS](#).

The example below shows a hypothetical country participating in TIMSS (both grades) and PIRLS. Students in the fourth grade sample will be administered both the TIMSS and PIRLS assessments, so that the fourth grade sample will serve for both. For reasons of administrative efficiency, the country wants to maximize the overlap between the fourth and eighth grade school samples. The 8,805 schools from the combined school frames (fourth and eighth grades) were first split in three strata (as described above). Then, a school sample of 164 was drawn the following way:

Example: Allocation of School samples in a Country Participating at Two Grade Levels

Overlap Strata	Total Sampled schools	Allocation		
		To PIRLS Grade 4	To TIMSS Grade 4	To TIMSS Grade 8
Grade 4 only	14	14	14	0
Grade 8 only	14	0	0	14
Grade 4 & Grade 8	136	136	136	136
Total	164	150	150	150

Choosing as many schools as possible from the “Grade 4 & Grade 8” stratum resulted in a sample of 150 schools (136+14) for each grade level, from a total of 164 sampled schools.

This sampling technique was most often used for TIMSS countries and benchmarking participants that had schools with students in both fourth and eighth grade populations. It was the preferred approach when the school MOS for fourth grade and eighth grade was well correlated, and when school samples could be drawn at the same time.

Sampling Method B:

Sampling Modifications for sequential data collection

This procedure could be used when Method A was not appropriate (low correlation between the school MOS for the samples, different sets of stratification variables for the samples, or school samples already drawn for one study – such as a national study involving the same school population). This procedure was used in Ireland, Sweden, and the benchmarking states in the United States.

Using this procedure, schools were sampled using a technique described in Chowdhury, Chu & Kaufman (2000). As explained by the authors, the method

can be used to either minimize or maximize overlap amongst several samples. This method is illustrated below with an example where the aim was to control overlap between a current sample of schools S_2 and a previously selected school sample S_1 . (For a complete description of the method, readers are referred to the original paper).

Let RL (Response Load) be the number of times a school was sampled from previous samples. In this example, given that there is only one previous sample, RL takes the value ‘1’ if the school was already selected and ‘0’ otherwise.

Given that the RL variable splits the current school frame in two distinct subsets of schools, S_1 and \bar{S}_1 , we have the following relation:

$$P_i(S_2) = P_i(S_2 | S_1) * P_i(S_1) + P_i(S_2 | \bar{S}_1) * P_i(\bar{S}_1) \quad (1)$$

where $P_i(S_j)$ gives the probability that school i be selected in the sample (S_j), and $P_i(S_j | S_k)$ gives the probability that school i be selected in sample (S_j) given that school i already belongs to (S_k). The idea here is to derive the conditional probabilities in such a way that the unconditional probability of selecting a school in the current sample, $P_i(S_2)$, be equal to the expected probability (as defined by the TIMSS sample design).

Note that the first term after the equal sign in equation (1) is related to cases where the school response load is one, while the last term is related to cases where the school response load is zero. Therefore, minimizing the sample overlap is equivalent to zeroing the first term. In such case, equation (1) becomes:

$$P_i(S_2) = 0 * P_i(S_1) + P_i(S_2 | \bar{S}_1) * P_i(\bar{S}_1) \quad \text{and consequently,}$$

$$P_i(S_2 | \bar{S}_1) = P_i(S_2) / P_i(\bar{S}_1)$$

In other words, in the current sample S_2 , schools would be selected with the following conditional probabilities:

0 if school i was already selected in the first sample,

$P_i(S_2) / P_i(\bar{S}_1)$ otherwise

However, equation (1) no longer holds if the expression $P_i(S_2)/P_i(\bar{S}_1)$ is greater than one. This can be avoided by setting one as an upper bound. We now have the following expression:

$$P_i(S_2) = P_i(S_2 | S_1) * P_i(S_1) + 1 * P_i(\bar{S}_1) \quad \text{and consequently}$$

$$\frac{P_i(S_2) - P_i(\bar{S}_1)}{P_i(S_1)} = P_i(S_2 | S_1)$$

Combining these two results, the conditional probabilities to use when selecting the current sample of schools are given by:

$$\begin{aligned} & \text{Max [} 0 \text{ , } \frac{P_i(S_2) - P_i(\bar{S}_1)}{P_i(S_1)} \text{]} \quad \text{if school } i \text{ was already selected} \\ & \hspace{15em} \text{in the first sample,} \\ & \text{Min [} P_i(S_2)/P_i(\bar{S}_1) \text{ , } 1 \text{]} \quad \text{otherwise} \end{aligned}$$

Note that maximizing rather than minimizing the overlap between two studies can be done by simply zeroing the last term of equation (1) rather than zeroing the first term, and following the above logic to get the conditional probabilities. The Chowdhury, Chu and Kaufman (2000) method can be generalized to more than two samples as described in their paper.

Further details about the implementation of this method for the countries and benchmark participants can be found in [TIMSS 2011 Characteristics of National Samples](#) and [PIRLS 2011 Characteristics of National Samples](#).

References

Chowdhury, S., Chu, A., & Kaufman, S. (2000). Minimizing overlap in NCES surveys. Proceedings of the Survey Methods Research Section, American Statistical Association, 174-179. Retrieved from http://www.amstat.org/sections/srms/proceedings/papers/2000_025.pdf