Appendix 3D: School Sampling Design Options to Accommodate Other Samples

TIMSS provides optional modifications to its sampling design for countries that want to maximize or minimize sampling overlap between schools sampled by TIMSS at the fourth and eighth grades as well as for countries that want to minimize overlap between schools sampled for TIMSS and schools sampled for other national or international assessments.

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 2015 with TIMSS at fourth grade and eighth grades) and the country wanted to control the overlap between the schools. Method B was used primarily to ensure that the TIMSS samples avoided schools sampled for other studies, and also used when Method A was not appropriate.

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 (i.e., fourth grade and eighth grade combined). Schools are sampled according to the sampling design described in this chapter. When selecting schools from strata comprising students from both populations, a country can choose to maximize or minimize the number of schools to be sampled at each grade level.

The example below shows a hypothetical country participating in TIMSS at both grades. 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 and then a school sample of 164 was drawn as shown in the accompanying table.

Method A: Allocation of School Samples in a Country Participating at Two Grade Levels

Overlap Strata	Total Sampled Schools	Allocation	
		To TIMSS Grade 4	To TIMSS Grade 8
Grade 4 only	14	14	0
Grade 8 only	14	0	14
Grade 4 & Grade 8	136	136	136
Total	**164**	**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. In this case, both studies were administered in the 136 schools selected from the Grade 4 & Grade 8 stratum.

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, where there was a strong correlation between the measure of size at both grades across these schools, and when school samples could be drawn at the same time.

Sampling Method B: Sampling Modifications for Sequential Data Collection

Method B was used to minimize overlap with another study such as a national study that also samples schools, and was also used when Method A was not appropriate (e.g., low correlation between MoS for fourth grade and eighth grade, samples not drawn simultaneously). In Method B, schools were sampled using a technique described in Chowdhury, Chu, and 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 minimize the 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 $\overline{S}_1$, we have the following relation:

$$P_i(S_2) = P_i(S_2|S_1) \cdot P_i(S_1) + P_i(S_2|\overline{S}_1) \cdot P_i(\overline{S}_1) \tag{1}$$

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:

$$P_i(S_2) = 0 \cdot P_i(S_1) + P_i(S_2|\overline{S}_1) \cdot P_i(\overline{S}_1)$$

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 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:

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

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:

$$\mathrm{Max}\left[0, \ \frac{P_i(S_2) - P_i(\bar{S}_1)}{P_i(S_1)}\right] \quad \text{if school } i \text{ was already selected in the first sample}$$

$$\mathrm{Min}\left[\frac{P_i(S_2)}{P_i(\bar{S}_1)}, \ 1\right] \quad \text{otherwise}$$

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 the Sample Implementation in TIMSS 2015 chapter.