

Chapter 5

TIMSS 2003 Sampling Design

Pierre Foy and Marc Joncas

5.1 Overview

This chapter describes the TIMSS 2003 international sample design and the procedures developed to ensure effective and efficient sampling of the student populations in each participating country. To be acceptable for TIMSS 2003, national sample designs had to result in probability samples that gave accurate weighted estimates of population parameters such as means and percentages, and for which estimates of sampling variance could be computed. The TIMSS 2003 sample design is similar to that used in TIMSS 1999, with minor refinements. Since sampling for TIMSS was to be implemented by the National Research Coordinator (NRC) in each participating country – often with limited resources – it was essential that the design be simple and easy to implement while yielding accurate and efficient samples of both schools and students. The design that was chosen for TIMSS strikes a good balance, providing accurate sample statistics while keeping the survey simple enough for all participants to implement.

The international project team provided software, manuals, and expert advice to help NRCs adapt the TIMSS sample design to their national system, and to guide them through the phases of sampling. The *School Sampling Manual* (TIMSS, 2001) describes how to implement the international sample design and to select the school sample; and offers advice on initial planning, adapting the design to national situations, establishing appropriate sample selection procedures, and conducting fieldwork. The *Survey Operations Manual* (TIMSS, 2002a) and *School Coordinator Manual* (TIMSS, 2002b) provide information on sampling within schools, assigning assessment booklets and questionnaires to sampled students, and tracking respondents and non-respondents. To automate the rather complex within-school sampling

procedures, NRCs were provided with sampling software jointly developed by the IEA Data processing Center (DPC) and Statistics Canada, documented in the *Within School Sampling Software (WinW3S) Manual* (TIMSS, 2002c).

In addition to sampling manuals and software, expert support was made available to help NRCs with their sampling activities. Statistics Canada and the IEA Data Processing Center (in consultation with the TIMSS sampling referee) reviewed and approved the national sampling plans, sampling data, sampling frames, and sample implementation. Statistics Canada and the DPC also provided advice and support to NRCs at all stages of the sampling process, drawing national school samples for nearly all of the TIMSS participants.

Where the local situation required it, NRCs were permitted to adapt the sample design for their educational systems, using more sampling information, and more sophisticated designs and procedures, than the base design required. However, these solutions had to be approved by the TIMSS International Study Center (ISC) at Boston College, and by Statistics Canada.

5.2 TIMSS Target Populations

In IEA studies, the target population for all countries is known as the *international desired population*. TIMSS 2003 chose to study achievement in two target populations, and countries were free to participate in either population, or both. The international desired populations for TIMSS were the following:

- **Population 1:** All students enrolled in the upper of the two adjacent grades that contain the largest proportion of 9-year-olds at the time of testing. This grade level was intended to represent four years of schooling, counting from the first year of primary or elementary schooling, and was the fourth grade in most countries.
- **Population 2:** All students enrolled in the upper of the two adjacent grades that contain the largest proportion of 13-year-olds at the time of testing. This grade level was intended to represent eight years of schooling, counting from the first year of primary or elementary schooling, and was the eighth grade in most countries.

To measure trends in student achievement, the TIMSS 2003 eighth- and fourth-grade target populations were intended to correspond to the upper grades of the TIMSS 1995 population definitions, and the TIMSS 2003 eighth-grade target population to the eighth-grade population in TIMSS 1999.

5.2.1 Sampling from the Target Populations

TIMSS expected all participating countries to define their *national desired populations* to correspond as closely as possible to its definition of the international desired populations.

For example, if fourth grade was the upper of the two adjacent grades containing the greatest proportion of 9-year-olds in a particular country, then all fourth grade students in the country should constitute the national desired population for that country.

Although countries were expected to include all students in the target grade in their definition of the populations, sometimes they had to restrict their coverage. Lithuania, for example, collected data only about students in Lithuanian-speaking schools, so their national desired populations fell short of the international desired populations. Appendix A of the TIMSS 2003 international reports in mathematics and science documents such deviations from the international definition of the TIMSS target populations.

Using their national desired populations as a basis, each participating country had to define its populations in operational terms for sampling purposes. This definition, known in IEA terminology as the national defined population, is essentially the sampling frame from which the first stage of sampling takes place. Ideally, the national defined populations should coincide with the *national desired populations*, although in reality there may be some school types or regions that cannot be included. Consequently, the national defined populations are usually a very large subset of the national desired populations. All schools and students in the desired populations not included in the defined populations are referred to as the excluded populations.

TIMSS participants were expected to ensure that the national defined populations included at least 95 percent of the national desired populations. Exclusions (which had to be kept to a minimum) could occur at the school level, within the sampled schools, or both. Because the national desired populations were restricted to schools that contained the required grade, schools not containing the target grade were considered to be outside the scope of the sample, i.e., not part of the target populations.

Although countries were expected to do everything possible to maximize coverage of the populations by the sampling plan, if necessary, schools could be excluded from the sampling frame for the following reasons:

- They were in geographically remote regions.
- They were of extremely small size.

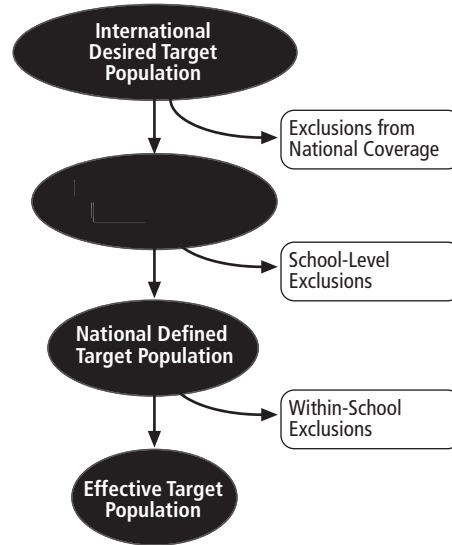
- They offered a curriculum or a school structure that was different from the mainstream education system(s).
- They provided instruction only to students in the categories defined as “within-school exclusions”.

Within-school exclusions were limited to students who, because of some disability, were unable to take part in the TIMSS assessment. The general TIMSS rules for defining within-school exclusions included the following three groups:

- **Intellectually disabled students.** These are students who were considered, in the professional opinion of the school principal or other qualified staff members, to be intellectually disabled, or who had been so diagnosed in psychological tests. This category included students who were emotionally or mentally unable to follow even the general instructions of the TIMSS tests. It did not include students who merely exhibited poor academic performance or discipline problems.
- **Functionally disabled students.** These are students who were permanently physically disabled in such a way that they could not perform on the TIMSS tests. Functionally disabled students who could perform were included in the testing.
- **Non-native language speakers.** These are students who could not read or speak the language of the test, and so could not overcome the language barrier of testing. Typically, a student who had received less than one year of instruction in the language of the test was excluded, but this definition was adapted in different countries.

Because these categories can vary internationally in the way they are implemented, NRCs were asked to adapt them to local usage. In addition, they were to estimate the size of the target population so that their compliance with the 95 percent rule could be projected. A major objective of TIMSS was that the effective target populations, the populations actually sampled by TIMSS, be as close as possible to the international desired populations. Exhibit 5.1 illustrates the relationship between the desired populations and the excluded populations. Each country had to account for any exclusion of eligible students from the international desired populations. This applied to school-level exclusions, as well as within-school exclusions.

Exhibit 5.1 Relationship Between the Desired Populations and Exclusions



5.3 Sample Design

The international sample design for TIMSS is generally referred to as a two-stage¹ stratified cluster sample design. The first stage consists of a sample of schools,² which may be stratified; the second stage consists of a sample of one or more classrooms from the target grade in sampled schools.

5.3.1 Units of Analysis and Sampling Units

The TIMSS analytical focus was on the cumulative learning of students, as well as on instructional characteristics related to learning. The sample design, therefore, had to address the measurement both of characteristics thought to influence cumulative learning, and of those specific to the instructional settings. As a consequence, although students were the principal units of analysis, schools and classrooms also were potential units of analysis, and all had to be considered as sampling units in the sample design in order to meet specific requirements for data quality and sampling precision at all levels.

Although the second stage sampling units were generally intact classrooms, the ultimate sampling elements were students – making it important that each student from a target grade be a member of one (and only one) of the classes in a school from which the sampled classes would be selected.

TIMSS prefers to sample intact classrooms because that allows the simplest link between students and teachers. In fourth grade, students in most countries are organized into classrooms that are taught as a unit for all

1 In some countries, it was necessary to include a third stage, where students within large classrooms were sub-sampled (see section 5.6).

2 In the Russian Federation, it was necessary to include an extra preliminary stage, where geographical regions were sampled first, and then schools (see section 5.4.3).

subjects, usually by the same teacher. Sampling intact classrooms is straightforward, therefore, at fourth grade. At eighth grade, however, classrooms are usually organized by subject – mathematics, language, science, etc. – and it is more difficult to arrange classroom sampling. TIMSS has addressed this issue by choosing the mathematics class as the sampling unit, mainly because classes often are organized on the basis of mathematics instruction and because mathematics is a central focus of the study. Although this is the recommended procedure, it can only be implemented where the mathematics classes in a school constitute an exhaustive and mutually exclusive partition of the students in the grade. This is the case when every student in the target grade attends one and only one mathematics class in the school.

5.3.2 Sampling Precision and Sample Size

In planning the sample design for each country, sample sizes for the two stages of the TIMSS sample design had to be specified so as to meet the sampling precision requirements of the study. Since students were the principal units of analysis, the reliability of estimates of student characteristics was paramount. However, TIMSS planned to report extensively on school, teacher, and classroom characteristics, so it was necessary also to have sufficiently large samples of schools and classes. The TIMSS standard for sampling precision requires that all student samples have an effective sample size of at least 400 students for the main criterion variables – mathematics and science achievement. In other words, all student samples should yield sampling errors that are no greater than would be obtained from a simple random sample of 400 students.

An effective sample size of 400 students results in the following approximate 95 percent confidence limits for sample estimates of population means, percentages, and correlation coefficients.

- Means: $m \pm 0.1s$ (where m is the mean estimate, and s is the estimated standard deviation for students)
- Percentages: $p \pm 5\%$ (where p is a percentage estimate)
- Correlations: $r \pm 0.1$ (where r is a correlation estimate)

Notwithstanding these precision requirements, TIMSS required a minimum of 4,000 students for each target population. This was necessary to ensure adequate sample sizes for sub-groups of students categorized by school, class, teacher, or student characteristics. Furthermore, since TIMSS planned to conduct analyses at the school and classroom levels, at least 150 schools were to be selected from each target population. Samples of 150 schools yield 95 percent confidence limits for school-level and classroom-

level mean estimates that are precise to within 16 percent of their standard deviations. Therefore, to ensure sufficient sample precision for school-level and student-level analyses, some participants had to sample more schools and students than would have been selected otherwise.

5.3.3 Clustering Effect

The precision of multistage cluster sample designs is generally affected by the so-called clustering effect. Students are clustered in schools, and are also clustered in classrooms within the schools. A classroom – as a sampling unit – constitutes a cluster of students who tend to be more like each other than like other members of the population. The intra-class correlation is a measure of this within-class similarity. Sampling 30 students from a single classroom when the intra-class correlation is high will yield less information than a random sample of 30 students drawn from across all students in the grade level. Consequently, a cluster sample with a positive intra-class correlation will need to have more elements than a random sample of independent elements to achieve the same level of precision. Thus, cluster sample designs are less efficient, in terms of sampling precision, than a simple random sample of the same size. This clustering effect was considered in determining the overall sample sizes for TIMSS.

The size of the cluster (classroom) and the size of the intra-class correlation determine the magnitude of the clustering effect. For planning its sample size, therefore, each country had to identify a value for the intra-class correlation and a value for the expected cluster size (this was known as the minimum cluster size). The intra-class correlation for each country was estimated from previous cycles of TIMSS, from IEA's Progress in International Reading Literacy Study (PIRLS), or from national assessments. In the absence of these sources, an intra-class correlation of 0.3 was assumed. Since participants were generally sampling intact classrooms, the minimum cluster size was in fact the average classroom size.

Sample-design tables, such as the one in Exhibit 5.2, were produced and included in the TIMSS *School Sampling Manual*. These tables illustrate the number of schools necessary to meet the TIMSS sampling precision requirements for a range of values of intra-class correlations and minimum cluster sizes. TIMSS participants could refer to the tables to determine how many schools they should sample. For example, on the basis of Exhibit 5.2, a participant whose intra-class correlation was expected to be 0.6, with an average classroom size of 30, would need to sample a minimum of 262 schools. Whenever the estimated number of schools to sample was less than 150, participants were asked to sample at least 150 schools. Also, if the total expected

number of students was less than 4,000, participating countries were asked to select more schools, or more classrooms per school. The sample design tables could also be used to determine sample sizes for more complex designs. For example, geographical regions could be defined as strata, whereby equal numbers of schools would be sampled in each stratum in order to produce equally reliable estimates for all strata, regardless of the relative size of the strata.

Exhibit 5.2 TIMSS Sample Design Table

MCS		Intraclass Correlation								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	a	212	244	276	308	340	372	404	436	468
	n	1,060	1,220	1,380	1,540	1,700	1,860	2,020	2,180	2,340
10	a	150	162	198	234	270	306	342	378	414
	n	1,500	1,620	1,980	2,340	2,700	3,060	3,420	3,780	4,140
15	a	150	150	172	209	247	284	321	359	396
	n	2,250	2,250	2,580	3,135	3,705	4,260	4,815	5,385	5,940
20	a	150	150	159	197	235	273	311	349	387
	n	3,000	3,000	3,180	3,940	4,700	5,460	6,220	6,980	7,740
25	a	150	150	151	190	228	266	305	343	382
	n	3,750	3,750	3,775	4,750	5,700	6,650	7,625	8,575	9,550
30	a	150	150	150	185	223	262	301	339	378
	n	4,500	4,500	4,500	5,550	6,690	7,860	9,030	10,170	11,340
35	a	150	150	150	181	220	259	298	337	375
	n	5,250	5,250	5,250	6,335	7,700	9,065	10,430	11,795	13,125
40	a	150	150	150	179	218	257	296	335	374
	n	6,000	6,000	6,000	7,160	8,720	10,280	11,840	13,400	14,960
45	a	150	150	150	176	216	255	294	333	372
	n	6,750	6,750	6,750	7,920	9,720	11,475	13,230	14,985	16,740
50	a	150	150	150	175	214	253	292	332	371
	n	7,500	7,500	7,500	8,750	10,700	12,650	14,600	16,600	18,550
55	a	150	150	150	173	213	252	291	331	370
	n	8,250	8,250	8,250	9,515	11,715	13,860	16,005	18,205	20,350
60	a	150	150	150	172	212	251	290	330	369
	n	9,000	9,000	9,000	10,320	12,720	15,060	17,400	19,800	22,140

a = Number of sampled schools

n = Number of sampled students in the target grade

Note: The Minimum Cluster Size (MCS) is the number of students selected in each sampled school (generally the average classroom size).

5.3.4 Stratification

Stratification is the grouping of sampling units (e.g., schools) in the sampling frame according to some attribute or variable prior to drawing the sample. It is generally used for the following reasons:

- To improve the efficiency of the sample design, thereby making survey estimates more reliable.
- To apply different sample designs or disproportionate sample-size allocations to specific groups of schools (such as those within certain states or provinces).
- To ensure adequate representation in the sample of specific groups from the target population.

Examples of stratification variables for school samples are: geography (such as states or provinces), school type (such as public and private), and level of urbanization (such as rural and urban). Stratification variables in the TIMSS sample design could be used explicitly, implicitly, or both.

- **Explicit stratification** consists of building separate school lists, or sampling frames, according to the stratification variables under consideration. For example, where geographic regions are an explicit stratification variable, separate school sampling frames would be constructed for each region. Different sample designs, or different sampling fractions, would then be applied to each school sampling frame to select the sample of schools. In TIMSS, the main reason for considering explicit stratification was to ensure disproportionate allocation of the school sample across strata. For example, a country stratifying by school type might require a specific number of schools from each stratum, regardless of the relative sizes of the strata.
- **Implicit stratification** makes use of a single school sampling frame, but sorts the schools in this frame by a set of stratification variables. This type of stratification, combined with the PPS systematic sampling methodology (see section 5.4), is a simple way of ensuring proportional sample allocation without the complexity of explicit stratification. It can also improve the reliability of survey estimates – provided the stratification variables are related to school mean student achievement in either mathematics or science.

5.3.5 Replacement Schools

Although TIMSS participants were expected to make great efforts to secure the participation of sampled schools, it was anticipated that a 100 percent participation rate would not be possible in all countries. To avoid sample size losses, a mechanism was instituted to identify, a priori, replacement schools for each sampled school. For each sampled school, the next school on the

ordered school sampling frame was identified as its replacement, and the one after that as a second replacement, should it be needed (see Exhibit 5.3 for an example).

The use of implicit stratification variables and the subsequent ordering of the school sampling frame by size ensured that any sampled school's replacement would have similar characteristics. Although this approach avoids sample size losses, it does not guarantee avoiding response bias. However, it may reduce the potential for bias, and was deemed more acceptable than over-sampling to accommodate a low response rate.

5.4 First Sampling Stage

The sample selection method used for the first sampling stage in TIMSS makes use of a systematic probability-proportional-to-size (PPS) technique. In order to use this method, it is necessary to have some measure of the size (MOS) of the sampling units. Ideally, this should be the number of sampling elements within the unit (e.g., the number of students in the school in the target grade). If this is unavailable, some other highly correlated measure, such as total school enrollment, may be used.

The schools in each explicit stratum are listed in order of the implicit stratification variables, together with the MOS for each school. Schools are further sorted by MOS within the implicit stratification variables. The measures of sizes are accumulated from school to school, and the running total (the cumulative MOS) is listed next to each school (see Exhibit 5.3). The cumulative MOS is an index of the size of the population of sampling elements; dividing it by the number of schools to be sampled gives the sampling interval.

The first school is sampled by choosing a random number in the range between 0 and the sampling interval. The school whose cumulative MOS contains the random number is the sampled school. By adding the sampling interval to that first random number, the second school is identified. This process of consistently adding the sampling interval to the previous selection number results in a PPS sample of schools of the required size.

Among the many benefits of this sample selection method are that it is easy to implement, and that it is easy to verify that it was implemented properly. The latter is critical, since one of the main methodological objectives of TIMSS was to ensure that a sound sampling technique had been used. Exhibit 5.3 illustrates the PPS systematic sampling method applied to a fictitious sampling frame. The first three sampled schools are shown, as well as their pre-selected replacement schools, which may be used should the originally selected schools not participate.

Exhibit 5.3 Application of the PPS Systematic Sampling Method to TIMSS

Total MOS: 392 154		Sampling Interval: 2 614.3600	
School Sample: 150		Random Start: 1 135.1551	
School Code	School MOS	Cumulative MOS	Sample
939438	532	532	
026825	517	1049	
277618	487	1536	–
228882	461	1997	R1
833389	459	2456	R2
386017	437	2893	
986694	406	3299	
041733	385	3684	
056595	350	4034	–
945801	341	4375	R1
865982	328	4703	R2
700089	311	5014	
656616	299	5313	
647690	275	5588	
381836	266	5854	
510529	247	6101	
729813	215	6316	
294281	195	6511	–
016174	174	6685	R1
292526	152	6837	R2
541397	133	6970	
502014	121	7091	
662598	107	7198	
821732	103	7301	
436600	97	7398	

– = Sampled School R1, R2 = Replacement Schools

5.4.1 Small Schools

Small schools, those with fewer eligible students than are typically found in a classroom, can cause difficulties in PPS sampling because students sampled from them tend to be assigned very large sampling weights, which can increase sampling variance. Also, because such schools supply fewer students than the other schools, the overall student sample size may be reduced. In TIMSS, a school was deemed to be small if the number of students in the target grade was less than the minimum cluster size. For example, if the minimum cluster size was set at 20, then a school with fewer than 20 students in the target grade was considered a small school.

The TIMSS approach for dealing with small schools had two components:

- **Exclude extremely small schools.** Extremely small schools were defined as schools with fewer students than one quarter of the minimum cluster size. For example, if the minimum cluster size was set at 20, schools with fewer than five students in the target grade were considered extremely small schools. If student enrollment in these schools was less than two percent of the eligible population, these schools could be excluded, provided the overall inclusion rate met the 95 percent criterion (see section 5.2.1).
- **Select remaining small schools with equal probabilities.** All remaining small schools were selected with equal probabilities within explicit strata. This was done by calculating, for each explicit stratum, the average size of small schools and setting the MOS of all small schools to this average size. The number of small schools to be sampled within explicit strata would thus remain proportional, and this action would ensure greater stability in the resulting sampling weights.

5.4.2 Very Large Schools

A very large school is a school whose measure of size is larger than the calculated sampling interval. Very large schools can cause operational problems because they stand a chance of being selected more than once under the normal PPS sampling method. This problem was solved in one of two ways:

- **Creating an explicit stratum of very large schools.** All very large schools were put in an explicit stratum and all of them were included in the sample. This was done within the originally defined explicit strata since the sampling intervals were calculated independently for each original explicit stratum. Thus, an explicit stratum would be divided into two parts if it contained any very large schools.
- **Setting their MOS equal to the sampling interval.** All very large schools in an explicit stratum were given a measure of size equal to the sampling interval calculated for that explicit stratum. In this way, very large schools were all included in the sample with probabilities of unity. This approach was simpler to apply and avoided the formation of additional explicit strata.

5.4.3 Optional Preliminary Sampling Stage

In TIMSS, very large countries have the opportunity to introduce a preliminary sampling stage before sampling schools. This consists of first drawing a sample of geographic regions using PPS sampling and then a sample of schools from each sampled region. This design is used mostly as a cost reduction

measure, where the construction of a comprehensive list of schools is either impossible or prohibitively expensive. Also, the additional sampling stage reduces the dispersion of the school sample, thereby potentially reducing travel costs. Sampling guidelines ensure that an adequate number of units are sampled from this preliminary stage. The sampling frame has to consist of at least 80 primary sampling units, of which at least 40 must be sampled at this stage. The Russian Federation was the only country to avail of this option in TIMSS 2003.

5.5 Second Sampling Stage

The second sampling stage in the TIMSS international design consisted of selecting classrooms within sampled schools. As a rule, one classroom per school was sampled, although some participants opted to sample two classrooms. Additionally, some participants were required to sample two or more classrooms per school in order to meet the minimum requirement of 4,000 sampled students. Classrooms were generally selected with equal probabilities. For those countries that chose to sub-sample students within classrooms (see section 5.6), classroom sampling was done using PPS sampling within the affected schools.

5.5.1 Small Classrooms

Generally, classrooms in an education system tend to be of roughly equal size. Occasionally, however, small classrooms are devoted to special situations, such as remedial or accelerated programs. These classrooms can become problematic in sampling, since they can lead to a shortfall in sample size, and also introduce some instability in the resulting sampling weights.

In order to avoid these problems, any classroom smaller than half the specified minimum cluster size was combined with another classroom from the same grade and school. For example, if the minimum cluster size was set at 30, any classroom with fewer than 15 students was combined with another. The resulting pseudo-classroom then constituted a sampling unit.

5.6 Sampling Students Within Classes

As a rule, all students in the sampled classrooms were expected to take part in the TIMSS assessment. However, countries where especially large classes were the norm could with permission opt to sub-sample a fixed number of students from each sampled classroom. Where applicable, this was done using a systematic sampling method whereby all students in a sampled classroom were assigned equal selection probabilities. In TIMSS 2003, only Yemen chose this option.

References

TIMSS (2001), *TIMSS 2003 School Sampling Manual*, prepared by P. Foy & M. Joncas, Statistics Canada, Chestnut Hill, MA: Boston College.

TIMSS (2002a), *TIMSS 2003 Survey Operations Manual*, prepared by the International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002b), *TIMSS 2003 School Coordinator Manual*, prepared by the International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002c), *TIMSS 2003 Within School Sampling Software (WinW3S) Manual*, prepared by O. Neuschmidt, J. Pickel et al., IEA Data Processing Center, Chestnut Hill, MA: Boston College.