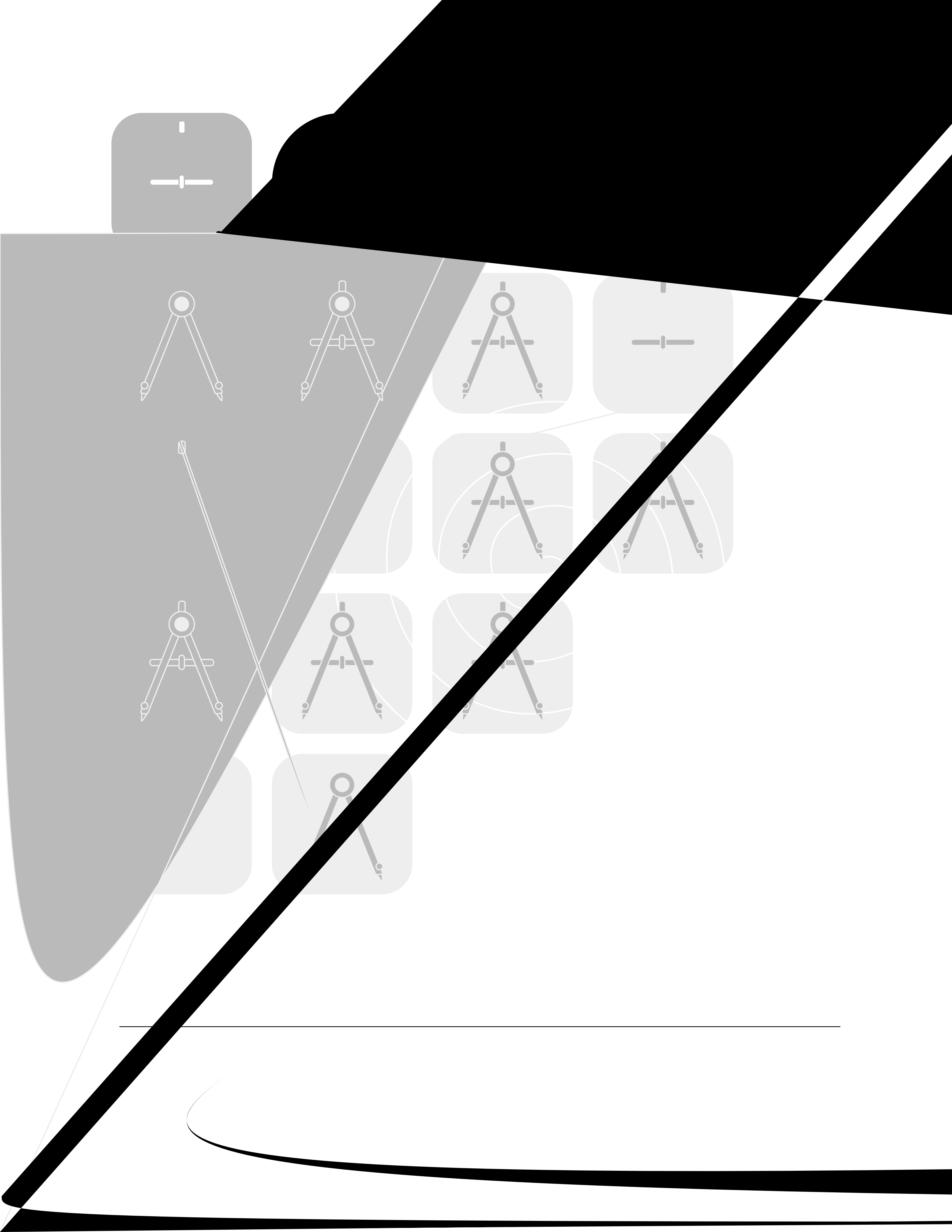
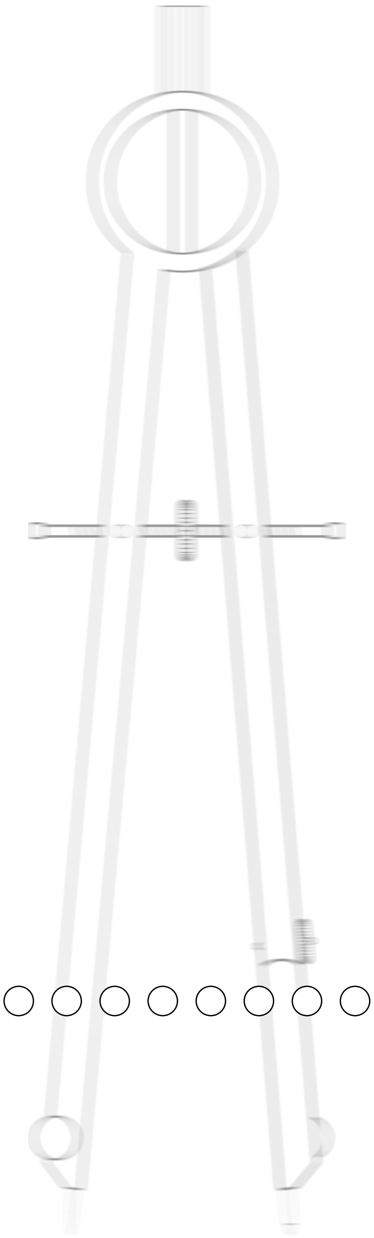




2 TIMSS Sample Design

Pierre Foy
Marc Joncas

2.1 Overview

This chapter describes the procedures developed to ensure proper sampling of the student populations in each participating country. To be acceptable for TIMSS 1999, national sample designs had to result in probability samples that gave accurately weighted estimates of population parameters, and for which estimates of sampling variance could be computed. The TIMSS 1999 sample design was very similar to that of its predecessor, TIMSS 1995, with minor refinements made as a result of the 1995 sampling. The TIMSS design was chosen so as to balance analytical requirements and operational constraints, while keeping it simple enough for all participants to implement. Representative and efficient samples in all countries were crucial to the success of the project. The quality of the samples depends on the sampling information available at the design stage, and particularly on the sampling procedures.

The National Research Coordinators (NRCs) were aware that in a study as ambitious as TIMSS 1999 the sample design and sampling procedures would be complex, and that gathering the required information about the national education systems would place considerable demands on resources and expertise. At the same time, those directing and coordinating the project realized that the national centers had only limited numbers of qualified sampling personnel. Keeping the procedures as simple as possible, especially the sample selection within schools, was thus a major consideration.

The international project management provided manuals and expert advice to help NRCs adapt the TIMSS 1999 sample design to their national system and to guide them through the phases of sampling. The TIMSS 1999 *School Sampling Manual* (TIMSS, 1997) described how to implement the international sample design and offered advice on planning, working within constraints, establishing appropriate sample selection procedures, and fieldwork. The *Survey Operations Manual* (TIMSS, 1998a) and *School Coordinator Manual* (TIMSS, 1998b) discussed sample selection and execution within schools, the assignment of test booklets to selected students, and administration and monitoring procedures used to identify and track respondents and nonrespondents. NRCs also received software designed to automate the sometimes complex within-school sampling procedures.

In addition, NRCs had access to expert support. Statistics Canada, in consultation with the TIMSS 1999 sampling referee, reviewed and approved the national sampling plans, sampling data, sampling frames, and sample selection. Statistics Canada also assisted nearly half of the TIMSS 1999 participants in drawing national school samples.

NRCs were allowed to adapt the basic TIMSS sample design to the needs of their education system by using more sampling information or more sophisticated designs and procedures. These adjustments, however, had to be approved by the International Study Center at Boston College and monitored by Statistics Canada.

2.2 Target Populations and Exclusions

In IEA studies, the target population for all countries is known as the *international desired population*. The international desired population for TIMSS 1999 was as follows:

- 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.

The TIMSS 1999 target grade was the upper grade of the TIMSS 1995 population 2 definition[1] and was expected to be the eighth grade in most countries. This would allow countries participating in both TIMSS 1995 and TIMSS 1999 to establish a trend line of comparable achievement data.

2.2.1 School and Within-School Exclusions

TIMSS 1999 expected all participating countries to define their *national desired population* to correspond as closely as possible to its definition of the international desired population. Sometimes, however, NRCs had to make changes. For example, some countries had to restrict geographical coverage by excluding remote regions; or to exclude a segment of their education system. The international reports document any deviations from the international definition of the TIMSS 1999 target population.

○○○

[1] For the TIMSS 1995 Population definition, see Foy, Rust, & Schleicher (1996).

Using their national desired population as a basis, participating countries had to operationally define their population 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. The national defined population could be a subset of the national desired population. All schools and students from the former excluded from the latter are referred to as the *excluded population.*

TIMSS 1999 participants were expected to keep the excluded population to no more than 10% of the national desired population. Exclusions could occur at the school level, within schools, or both. Because the national desired population was restricted to schools that contained the target grade, schools not containing this grade were considered to be outside the scope of the sampling frame, and not part of the excluded population. Participants could exclude schools 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 school structure, that was different from the mainstream education system(s).
- They provided instruction only to students in the exclusion categories defined as "within-sample exclusions."

Within-sample exclusions were limited to students who, because of some disability, were unable to take the TIMSS 1999 tests. NRCs were asked to define anticipated within-sample exclusions. Because these definitions can vary internationally, NRC's were also asked to follow certain rules adapted to their jurisdictions. In addition, they were to estimate the size of such exclusions so that compliance with the 10% rule could be gauged in advance.

The general TIMSS 1999 rules for defining within-school exclusions included:

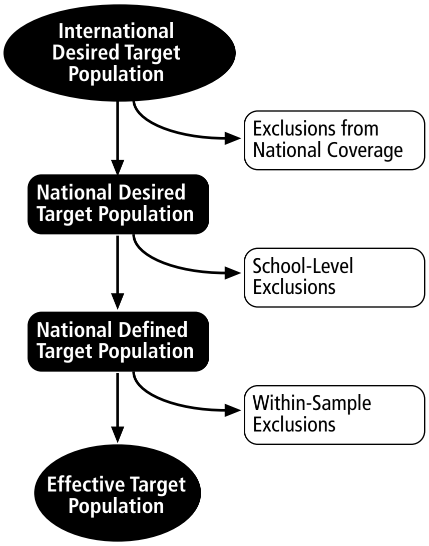
- **Educable mentally disabled students**. These are students who were considered, in the professional opinion of the school principal or other qualified staff members, to be educable mentally disabled, or students who had been so diagnosed by psychological tests. This included students who were emotionally or mentally unable to follow even the general instructions of the TIMSS 1999 test. 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 in the TIMSS 1999 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.

The stated objective in TIMSS 1999 was that the effective target population, the population actually sampled by TIMSS 1999, be as close as possible to the international desired population. Exhibit 2.1 illustrates the relationship between the desired populations and the excluded populations. Any exclusion of eligible students from the international desired population had to be accounted for, both at the school level and within samples.

The size of the excluded population was documented and served as an index of the coverage and representativeness of the selected samples.

Exhibit 2.1 Relationship Between the Desired Populations and Exclusions


International Desired Target Population
Exclusions from National Coverage
National Desired Target Population
School-Level Exclusions
National Defined Target Population
Within-Sample Exclusions
Effective Target Population

2.3 Sample Design

The basic sample design for TIMSS 1999 is generally referred to as a two-stage stratified cluster sample design. The first stage consisted of a sample of schools[2], which may be stratified; the second stage consisted of a single mathematics classroom selected at random from the target grade in sampled schools. It was also permissible to add a third stage, in which students could be sampled within classrooms. This design lent itself to the many analytical requirements of TIMSS 1999.

2.3.1 Units of Analysis and Sampling Units

The TIMSS 1999 analytical focus was both on the cumulative learning of students and on the instructional characteristics affecting learning. The sample design, therefore, had to address the measurement both of characteristics thought to influence cumulative learning and of specific characteristics of instruction. Because schools, classrooms, and students were all considered potential units of analysis, they had to be considered as sampling units. This was necessary in order to meet specific requirements for data quality and sampling precision at all levels.

[2] In some very large countries, it was necessary to include an extra preliminary stage in which school districts were sampled first, and then schools.

Although in the second sampling stage the sampling units were intact mathematics classrooms, the ultimate sampling elements were students. Consequently, it was important that each student from the target grade be a member of one and only one of the mathematics classes in a school from which the sampled classes were to be selected. In most education systems, the mathematics class coincided with a student homeroom or science class. In some systems, however, mathematics and science classes did not coincide. In any case, participating countries were asked to define the classrooms on the basis of mathematics instruction. If not all students in the national desired population belonged to a mathematics class, then an alternative definition of the classroom was required for ensuring that the non-mathematics students had an opportunity to be selected.

2.3.2 Sampling Precision and Sample Size

Sample sizes for TIMSS 1999 had to be specified so as to meet the analytic requirements of the study. Since students were the principal units of analysis, the ability to produce reliable estimates of student characteristics was important. The TIMSS 1999 standard for sampling precision required that all population samples have an effective sample size of at least 400 students for mathematics and science achievement. In other words, the samples should have sampling errors no greater than those that would be obtained from a simple random sample of 400 students.

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

- Means: m ± 0.1s (where m is the mean estimate and s is the estimated standard deviation for students)
- Percentages: p ± 5.0% (where p is a percentage estimate)
- Correlations: r ± 0.1 (where r is a correlation estimate)

Furthermore, since TIMSS 1999 was designed to allow for analyses at the school and classroom levels, at least 150 schools were to be selected from the target population. A sample of 150 schools results in 95% confidence limits for school-level and classroom-level mean estimates that are precise to within ± 16% of their standard deviations. To ensure sufficient sample precision for these units of analysis, some participants had to sample more schools than they would have selected otherwise.

The precision of multistage cluster sample designs are generally affected by the so-called clustering effect. 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 *intraclass correlation* is a measure of this similarity. Sampling 30 students from a single classroom, when the intraclass correlation is positive, will yield less information than a random sample of 30 students spread across all classrooms in a school. Such sample designs are less efficient, in terms of information per sampled student, than a simple random sample of the same size. This clustering effect had to be considered in determining the overall sample size for TIMSS 1999.

The magnitude of the clustering effect is determined by the size of the cluster (classroom) and the size of the intraclass correlation. For planning the sample size, therefore, each country had to choose a value for the intraclass correlation, and a value for the expected cluster size (this was known as the minimum cluster size). The intraclass correlation for each country was estimated from past studies, such as TIMSS 1995, or from national assessments. In the absence of such sources, an intraclass correlation of 0.3 was assumed. Since all participants chose to test intact classrooms, the minimum cluster size was in fact the average classroom size. The specification of the minimum cluster size affected not only the number of schools sampled, but also the way in which small schools and small classrooms were treated.

Sample-design tables were produced and included in the TIMSS 1999 School Sampling Manual (see Exhibit 2.2 for an example). These tables illustrated the number of schools that had to be sampled to meet the TIMSS sampling precision requirements for a range of values of intraclass correlation and minimum cluster sizes. TIMSS 1999 participants could use these tables to determine how many schools they should sample. For example, an examination of Exhibit 2.2 shows that a participant whose intraclass correlation was expected to be 0.6 and whose average classroom size was 30 needed to sample a minimum of 248 schools. Whenever the estimated number of schools to sample fell below 150, participants were asked to sample at least 150 schools.

The sample-design tables could be used also to determine sample sizes for more complex designs. For example, a number of strata could be constructed for which different minimum cluster sizes could be specified, thereby refining the national sample design in a way that might avoid special treatment of small schools (See section 2.3.6, Small Schools).

Exhibit 2.2: Sample-Design Table* (95%Confidence Limits For Means ±0.1s / Percentages ±5.0)

MCS**		Intraclass Correlation								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	a	150	157	189	221	253	285	317	349	381
	n	750	785	945	1 105	1 265	1 425	1 585	1 745	1 905
10	a	150	150	155	191	227	263	299	335	371
	n	1 500	1 500	1 550	1 910	2 270	2 630	2 990	3 350	3 710
15	a	150	150	150	180	218	255	292	330	367
	n	2 250	2 250	2 250	2 700	3 270	3 825	4 380	4 950	5 505
20	a	150	150	150	175	213	251	289	327	365
	n	3 000	3 000	3 000	3 500	4 260	5 020	5 780	6 540	7 300
25	a	150	150	150	172	211	249	287	326	364
	n	3 750	3 750	3 750	4 300	5 275	6 225	7 175	8 150	9 100
30	a	150	150	150	170	209	248	286	325	364
	n	4 500	4 500	4 500	5 100	6 270	7 440	8 580	9 750	10 920
		150	150	150	169	208	246	285	324	363
	[redacted]	[redacted]	5 250	5 250	5 915	7 280	8 610	93c313BT7 02 352.334 Tm0 0 0 sc(150)T.361 307 440285		

a = number of sampled schools

n = number of sampled students in target grade

*Minimum school sample required = 150

**MCS is the number of students selected in each sampled school (generally the average classroom size).

2.3.3 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 schools), and level of urbanization (such as rural and urban). Stratification variables in the TIMSS 1999 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. Where, for example, geographic regions were an explicit stratification variable, separate school sampling frames were constructed for each region. Different sample designs, or different sampling fractions, could then be applied to each school-sampling frame to select the sample of schools. In practice, the main reason for considering explicit stratification in TIMSS 1999 was disproportionate allocation of the school sample across strata. For example, a country might require an equal number of schools from each stratum, regardless of the relative size of each stratum.

Implicit stratification makes use of a single school sampling frame, but sorts the schools in this frame by a set of stratification variables. This is a simple way of ensuring proportional sample allocation without the complexity of explicit stratification. Implicit stratification can also improve the reliability of survey estimates, provided the variables are related to school mean student achievement in mathematics and science.

2.3.4 Replacement Schools

Although TIMSS participants placed great emphasis on securing school participation, it was anticipated that a 100% participation rate would not be possible in all countries. To avoid losses in sample size, a mechanism was instituted to identify, a priori, two replacement schools for each sampled school. 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 was not guaranteed to avoid response bias, it would tend to minimize the potential for bias. Furthermore, it was deemed more acceptable than over-sampling to accommodate a low response rate.

2.3.5 First Sampling Stage

The sample-selection method used for the first-stage of sampling in TIMSS 1999 made use of a systematic probability-proportional-to-size (PPS) technique. Use of this method required some measure of size (MOS) of the sampling units. Ideally this was the number of sampling elements within the unit (e.g., number of students in the target grade in the school). If this information was unavailable, some other highly correlated measure, such as total school enrollment, was used.

The schools in each explicit stratum were listed in order of the implicit stratification variables, together with the MOS for each school. They were further sorted by MOS within variable. The measures of size were accumulated from school to school, and the running total (the cumulative MOS) was listed next to each school (see Exhibit 2.3). The cumulative MOS was a measure of the size of the population of sampling elements; dividing it by the number of schools sampled gives the *sampling interval.*

The first school was sampled by choosing a random number in the range between 1 and the sampling interval. The school whose cumulative MOS contained the random number was the sampled school. By adding the sampling interval to that first random number, a second school was identified. This process of consistently adding the sampling interval to the previous selection number resulted in a PPS sample of the required size.

As each school was selected, the next school in the sampling frame was designated as a replacement school for use should the sampled school not participate in the study, and the next after that as a second replacement, for use should neither the sampled school nor its replacement participate.

Two of the many benefits of the PPS sample selection method are that it is easy to implement, and that it is easy to verify that it was implemented properly. The latter was critical since one of TIMSS 1999's major objectives was to be able to verify that a sound sampling methodology had been used.

Exhibit 2.3 illustrates the PPS systematic sampling method applied to a fictitious sampling frame. The first three sampled schools are shown, as well as their corresponding first and second replacements (R1 and R2).

Exhibit 2.3: Application of the PPS Systematic Sampling Method

Total MOS:	392154	Sampling Interval:	2614.3600
School Sample:	150	Random Start:	1135.1551

School Identification Number	Measure of Size (MOS)	Cumulative MOS	Sampled and Replacement Schools
172989	532	532	
976181	517	1049	
564880	487	1536	S
387970	461	1997	R1
483231	459	2456	R2
550766	437	2893	
228699	406	3299	
60318	385	3684	
201035	350	4034	S
107346	341	4375	R1
294968	328	4703	R2
677048	311	5014	
967590	299	5313	
644562	275	5588	
32562	266	5854	
194290	247	6101	
129135	215	6316	
1633	195	6511	S
256393	174	6685	R1
754196	152	6837	R2
750793	133	6970	
757843	121	7091	
743500	107	7198	
84930	103	7301	
410355	97	7398	

S = Sampled School
R1, R2 = Replacement Schools

2.3.6 Small Schools

Small schools tend to be problematic in PPS samples because students sampled from these schools get disproportionately large sampling weights, and when the school size falls below the minimum cluster size, it reduces the overall student sample size. A school was deemed small in TIMSS 1999 if it was smaller than the minimum cluster size. Thus, if the minimum cluster size for a country was set at 20, then a school with fewer than 20 students in the target grade was considered a small school.

In TIMSS 1999, small schools were handled differently than in TIMSS 1995. The 1999 approach for dealing with them consisted of two steps

- **Extremely small schools**. Extremely small schools were defined as schools with fewer students than half the minimum cluster size. For example, if the minimum cluster size was set at 20, then schools with fewer than 10 students in the target grade were considered extremely small schools. If student enrollment in these schools was less than 2% of the eligible population, they were excluded, provided the overall exclusion rate did not exceed the 5% criterion (see Section 2.3).
- **Explicit stratum of small schools**. If fewer than 10% of eligible students were enrolled in small schools, then no additional action was required. If, however, more than 10% of eligible students were enrolled in small schools, then an explicit stratum of small schools was required. The number of schools to sample from this stratum remained proportional to the stratum size, but all schools had an equal probability of selection. This action ensured greater stability in the resulting sampling weights.

2.3.7 Optional Preliminary Sampling Stage

Some very large countries chose to introduce a preliminary sampling stage before sampling schools. This consisted of a PPS sample of geographic regions. A sample of schools was then selected from each sampled region. This design was used mostly as a cost-reduction measure where the construction of a comprehensive list of schools would have been either impossible or prohibitively expensive. Also, this additional sampling stage reduced the dispersion of the school sample, thereby potentially reducing travel costs. Sampling guidelines were put in place to ensure that an adequate number of units were sampled from this preliminary stage. The sampling frame had to consist of at least 80 primary sampling units, of which at least 40 had to be sampled at this stage.

2.3.8 Second Sampling Stage

The second sampling stage consisted of selecting classrooms within sampled schools. As a rule, one classroom per school was sampled, although some participants opted to sample two classrooms. Classrooms were selected either with equal probabilities or with probabilities proportional to their size. Participants who opted to test all students in selected classrooms sampled classrooms with equal probabilities. This was the method of choice for most participants. A procedure was also available whereby NRCs could choose to sub-sample students within selected classrooms using PPS.

2.3.9 Small Classrooms

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

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

2.4 Participation Rates

Weighted and unweighted response rates were computed for each participating country at the school level and at the student level. The basic formulae for response rates are provided in this section. More elaborate treatment of participation rates, including adjustments for non-participation, may be found in Chapter 11.

2.4.1 School-Level Participation Rates

The minimum acceptable school-level participation rate, before the use of replacement schools, was set at 85%. This criterion was applied to the unweighted school response rate. School response rates were computed and reported both weighted and unweighted, with and without replacement schools. The general formula for computing weighted school-level response rates is shown in the following equation:

$$R_{wgt}(sch) = \frac{\sum_{part} MOS_i/\pi_i}{\sum_{elig} MOS_i/\pi_i}$$

For each sampled school, the ratio of its measure of size (MOS) to its selection probability (π_i) is computed. The weighted school-level participation rate is the sum of the ratios for all participating schools divided by the sum of the ratios for all eligible schools. The unweighted school-level participation rates are computed in a similar way, with all school ratios set to unity. This becomes simply the number of participating schools in the sample divided by the number of eligible schools in the sample. Since in most cases, in selecting the sample, the value of π_i was set proportional to MOS_i within each explicit stratum, weighted and unweighted rates are generally similar.

2.4.2 Student-Level Participation Rates

Like the school-level participation rate, the minimum acceptable student-within-school participation rate was set at 85%. This criterion was applied to the unweighted student-level participation rate. Both weighted and unweighted student participation rates were computed and reported. The general formula for computing student-level participation rates is shown in the following equation:

$$R_{wgt}(std) = \frac{\sum_{part} 1/p_j}{\sum_{elig} 1/p_j}$$

where p_j denotes the probability of selection of the student, incorporating all stages of selection. Thus the weighted student-level participation rate is the sum of the inverse of the selection probabilities for all participating students divided by the sum of the inverse of the selection probabilities for all eligible students. The unweighted student participation rates were computed in a similar way, but with each student contributing equal weight.

2.4.3 Overall Participation Rates

The minimum acceptable overall response rate was set at 75%. This rate was calculated as the product of the weighted school-level participation rate without replacement schools and the weighted student-level participation rate. Weighted overall participation rates were computed and reported both with and without replacement schools.

References

Foy, P., Rust, K., & Schleicher, A. (1996). Sample Design in M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report Volume I: Design and Development.* Chestnut Hill, MA: Boston College.

TIMSS (1997). *TIMSS 1999 School Sampling Manual–Version 2* (Doc. Ref.: TIMSS 1999 97-0012). Prepared by Pierre Foy, Statistics Canada. Chestnut Hill, MA: Boston College.

TIMSS (1998a). *Survey Operations Manual–Main Survey* (Doc. Ref.: TIMSS 1999 98-0026). Prepared by the International Study Center. Chestnut Hill, MA: Boston College.

TIMSS (1998b). *School Coordinator Manual–Main Survey* (Doc. Ref.: TIMSS 1999 98-0024). Prepared by the International Study Center. Chestnut Hill, MA: Boston College.