Chapter 4: SAMPLE DESIGN

TARGET POPULATION AND OVERVIEW OF THE SAMPLING DESIGN

The desired base PISA-D target population in each country consisted of 15-year-old students attending educational institutions in grades 7 and higher. This meant that countries were to include:

- 15-year-olds enrolled full-time in educational institutions;
- 15-year-olds enrolled in educational institutions who attended only on a part-time basis;
- students in vocational training programmes, or any other related type of educational programmes; and
- students attending foreign schools within the country (as well as students from other countries attending any of the programmes in the first three categories).

It was recognised that no testing of 15-year-olds schooled in the home, workplace, or out of the country would occur and therefore these out-of-school 15-year-olds were not included in the international target population.

The operational definition of an age population directly depends on the testing dates. The international requirement was that the assessment had to be conducted during a 42-day period, referred to as the testing period, between 1 September and 30 November 2017, unless otherwise agreed. There were three countries (Cambodia, Guatemala, and Senegal) whose testing periods were not confined to this window due to the academic calendar or other important operational considerations.

Further, testing was not permitted during the first six weeks of the school year due to concern that student performance levels may have been lower at the beginning of the academic year than at the end of the previous academic year, even when controlling for age.

The 15-year-old international target population for PISA was constructed to better fit the age structure of most of the participating Northern Hemisphere countries. As the majority of PISA testing has historically been planned to occur in April, the international target population was consequently defined as all students aged from 15 years and 3 completed months to 16 years and 2 completed months at the beginning of the assessment period. This generally allowed the target population to be defined by birth dates within a single, full calendar year (i.e. 1999). To maintain consistency with PISA, the international target population for PISA-D was also defined as students aged from 15 years and 3 months to 16 years and 2 months. However, since the testing was to take place during the later months of the calendar year, the 12-month period used to define each PISA-D country's eligible birth dates could not be confined to a single calendar year.

Also consistent with PISA, a variation of up to one month in this age definition was permitted. Therefore, a country's eligible students could be defined as those aged 15 years and 2 completed months to 16 years and 1 completed month, or those aged 15 years and 4 completed months to

16 years and 3 completed months. This definition was set prior to carrying out sampling for each country.

The sampling design used for PISA-D was a two-stage stratified sample design. The first-stage sampling units consisted of individual schools with PISA-D eligible students, or with the possibility of having such students at the time of assessment. Schools were sampled systematically from a comprehensive national list of all PISA-D-eligible schools, known as the school sampling frame, with probabilities that were proportional to a measure of size. The measure of size was a function of the estimated number of PISA-D-eligible 15-year-old students enrolled in the school. This is referred to as systematic Probability Proportional to Size (PPS) sampling. Prior to sampling, schools in the sampling frame were assigned to mutually exclusive groups based on school characteristics called explicit strata, formed to improve the precision of sample-based estimates. The characteristics of this assignment varied country by country.

The second-stage sampling units were students within sampled schools. Once schools were selected to be in the sample, a complete list of each sampled school's 15-year-old students was prepared. For each country a Target Cluster Size (TCS) was set; this value was typically 42 students, although with agreement countries could use alternative values. The sample size within schools was prescribed, within limits, in the PISA-D Technical Standards (see Annex D). From each list of students in a school that contained more than the TCS, a sample of typically 42 students was selected with equal probability, and for lists with fewer than the TCS, all students on the list were selected.

POPULATION COVERAGE, AND SCHOOL AND STUDENT PARTICIPATION RATE STANDARDS

To provide valid estimates of student achievement, the sample of students had to be selected using established and professionally recognised principles of scientific sampling in a way that ensured representation of the full target population of 15-year-old students in the participating countries.

Furthermore, quality standards had to be maintained with respect to (i) the coverage of the PISA-D international target population, (ii) accuracy and precision, and (iii) the school and student response rates.

Coverage of the PISA-D international target population

National Project Managers (NPMs) might have found it necessary to reduce their coverage of the target population by excluding, for instance, a small, remote geographical region due to inaccessibility, or a language group, possibly due to political, organisational, or operational reasons, or special education needs students. Areas deemed to be part of a country (for the purpose of PISA-D) but not included for sampling, although this occurred infrequently, were designated as non-covered areas. Care was taken in this regard because, when such situations did occur, the national desired target population differed from the international desired target population. In an international survey in education, the types of exclusion must be defined consistently for all participating countries and the exclusion rates have to be limited. Indeed, if a significant proportion of students were excluded, this would mean that survey results would not

be deemed representative of the entire national school system. Thus, efforts were made to ensure that exclusions, if they were necessary, were minimised according to the PISA-D Technical Standards.

Exclusion can also take place either at the school level (exclusion of entire schools), often for Special Education Needs (SEN) schools, or language schools, or at the within-school level (exclusion of individual students), most often for SEN needs or language. International within-school exclusion rules for students were specified as follows:

- Intellectually disabled students were students with a mental or emotional disability and who, in the professional opinion of qualified staff, were cognitively delayed such that they could not be validly assessed in the PISA-D testing setting. This category included students who were emotionally or mentally unable to follow even the general instructions of the test. Students were not to be excluded solely because of poor academic performance or normal discipline problems.
- Functionally disabled students were students who were permanently physically disabled in such a way that they cannot be validly assessed in the PISA-D testing setting. However, functionally disabled students who could provide responses were to be included in the testing.
- Students with insufficient assessment language experience were students who needed to meet all of the following criteria: i) were not native speakers of the assessment language(s), ii) had limited proficiency in the assessment language(s), and iii) had received less than one year of instruction in the assessment language(s). Students with insufficient assessment language experience could be excluded.
- Students not assessable for other reasons as agreed upon. A nationally defined within-school exclusion category was permitted if agreed upon by the international contractor. A specific subgroup of students (for example, students with severe dyslexia, dysgraphia, or dyscalculia) could be identified for whom exclusion was necessary but for whom the previous three within-school exclusion categories did not explicitly apply so that a more specific within-school exclusion definition was needed. No country defined not assessable for other reasons.
- Students taught in a language of instruction for the main domain for which no materials were available. Standard 2.1 notes that PISA-D was administered to a student in a language of instruction provided by the sampled school to that sampled student in the major domain of the test. Thus, if no test materials were available in the language in which the sampled student is taught, the student was excluded. For example, if a country had materials for testing in language X, but a sampled student was taught in language A, then this student taught in language A could be excluded since there were no materials available for that student to be tested in his or her language of instruction.

A school attended only by students who would be excluded for the reasons listed above (intellectual, functional, or linguistic reasons) was considered a school-level exclusion.

It was required that the overall exclusion rate within a country (i.e., school-level and within-school

exclusions combined) be kept below 5% of the PISA-D Desired Target Population. Guidelines for restrictions on the level of exclusions of various types were as follows:

- School-level exclusions for inaccessibility, feasibility or other reasons were to cover less than 0.5% of the total number of students in the PISA-D Desired Target Population for participating countries. Schools on the school sampling frame that had only one or two PISA-D-eligible students were not allowed to be excluded from the frame. However, if, based on the frame, it was clear that the percentage of students in these small schools would not cause a breach of the 0.5% allowable limit, then such schools could be excluded in the field at the time of the assessment if they still only had one or two PISA-D-eligible students.

- School-level exclusions for intellectually or functionally disabled students, or students with insufficient assessment language experience, were to cover fewer than 2% of the PISA-D Desired Target Population of students.

- Within-school exclusions for intellectually disabled or functionally disabled students, or students with insufficient assessment language experience, or students nationally-defined and agreed upon for exclusion were expected to cover fewer than 2.5% of PISA-D students. Initially, this could only be an estimate.

Accuracy and precision

A minimum of 150 schools had to be selected in each country. Within each participating school, a predetermined number of PISA-D eligible students, denoted as TCS (usually 42 students), were randomly selected with equal probability. In schools with fewer than TCS eligible students, all students were selected. In total, a minimum sample size of 5,250 assessed students was to be achieved. It was possible to negotiate a TCS that differed from 42 students, but if it was reduced, then the minimum sample size of schools was increased beyond 150 to ensure that at least the minimum required number of assessed students would be achieved. The TCS selected per school had to be at least 20 students to ensure adequate accuracy in estimating variance components within and between schools—a major analytical objective of PISA-D.

NPMs were strongly encouraged to identify available variables to use for defining the explicit and implicit strata for schools to reduce the sampling variance. See the section on stratification below for further details.

School response rates

Figure 4.1 provides a summary of the international requirements for school response rates. A response rate of 85% was required for initially selected schools and was achieved for all PISA-D countries. To compensate for a sampled school that did not participate, where possible, two potential replacement schools were identified. The school replacement process is described in the section below on "School Sample Selection".

Furthermore, a school with a student participation rate between 25% and 50% was not considered

as a participating school for the purposes of calculating and documenting response rates.[1] However, data from such schools were included in the database and contributed to the estimates included in the initial PISA-D international report. Data from schools with a student participation rate of less than 25% were not included in the database, and such schools were regarded as non-respondents.

PISA developed the following rationale for this approach. There was concern that, in an effort to meet the requirements for school response rates, a National Centre might accept participation from schools that would not make a concerted effort to have students attend the assessment sessions. To avoid this, a standard for student participation was required for each individual school in order that the school be regarded as a participant. This standard was set at a minimum of 50% student participation. However, there were a few schools in many countries that conducted the assessment without meeting that standard. Thus a judgement was needed to decide if the data from students in such schools should be used in the analyses given that the students had already been assessed. If the students from such schools were retained, non-response bias would possibly be introduced to the extent that the students who were absent could have been different in achievement from those who attended the testing session, and such a bias would be magnified by the relative sizes of these two groups. If one chose to delete all assessment data from such schools, then non-response bias would be introduced to the extent that the school was different from others in the sample, and sampling variance would be increased because of sample size attrition.

Therefore, the judgement was made that, for a school with between 25% and 50% participating students, the latter source of bias and variance was likely to introduce more error into the study estimates than the former. But the converse judgement made sense for those schools with a student response rate below 25%. Clearly the cut-off of 25% was arbitrary as one would need extensive studies to try to establish this cut-off empirically. However, it is clear that, as the student response rate decreases within a school, the possibility of bias from using the assessed students in that school will increase, while the loss in sample size from dropping all of the students in the school will be small.

---

[1]   Students were deemed participants if they responded to at least half of the cognitive items, or if they had responded to at least one cognitive item and had completed the background questionnaire (see Annex D).

■ Figure 4.1 ■

School response rate standards
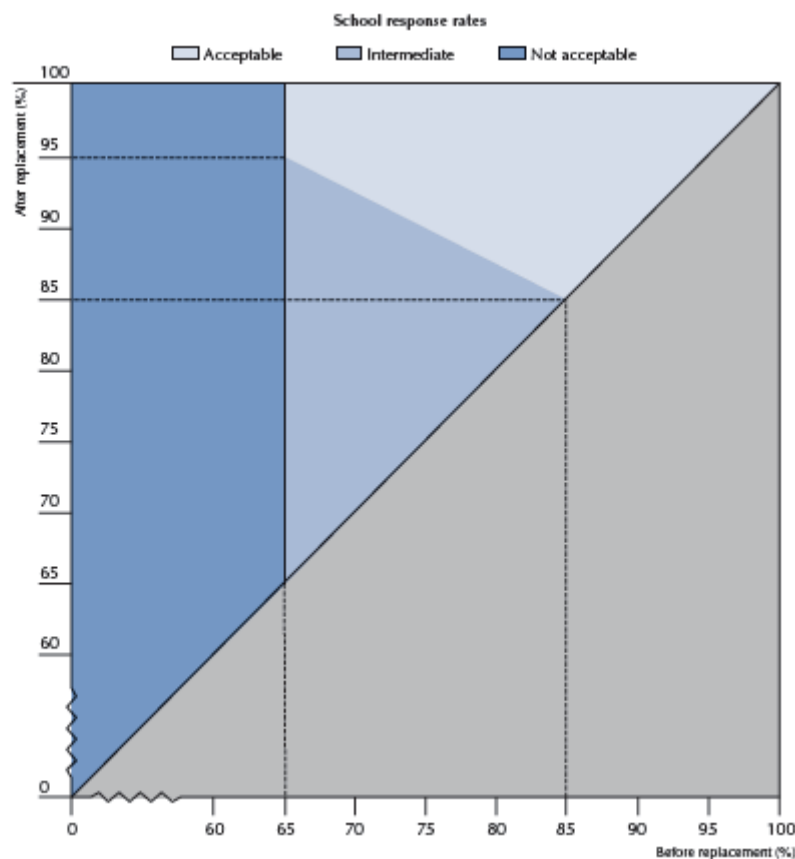
These PISA-D standards applied to weighted school response rates. The procedures for calculating weighted response rates are presented in Chapter 8. Weighted response rates weight each school by the number of students in the population that are represented by the students sampled from within that school. The weight consists primarily of the enrolment size of 15-year-old students in the school divided by the selection probability of the school. Because the school samples were selected with probability proportional to size, in most countries most schools contributed approximately equal weights. As a consequence, the weighted and unweighted school response rates were similar. Details as to how each participating country performed relative to these school response rate standards are included in Chapters 11 and 14.

Student response rates

An overall response rate of 80% of selected students in participating schools was required. A student who had participated in the original or follow-up cognitive sessions was considered to be a participant. A minimum student response rate of 50% within each school was required for a school to be regarded as participating: the overall student response rate was computed using only students from schools with at least a 50% student response rate. Again, weighted student response rates were used for assessing this standard. Each student was weighted by the reciprocal of his or her sample selection probability.

MAIN SURVEY SCHOOL SAMPLE

Definition of the national target population

NPMs were first required to confirm their dates of testing and age definition with the International contractor. Once these were approved, NPMs were alerted to avoid having the possible drift in the assessment period lead to an unapproved definition of the national target population.

Every NPM was required to provide counts of his or her country's target population and explain how and why it might deviate from the international target population. Any hardships in accomplishing complete coverage were specified, discussed, and approved or not, in advance. Where the national target population deviated from full coverage of all PISA-D-eligible students, the deviations were described and enrolment data provided to measure how much coverage was reduced. The population, after all exclusions, corresponded to the population of students recorded on each country's school sampling frame. Exclusions were often proposed for practical reasons such as increased survey costs or complexity in the sample design and/or difficult test conditions. These difficulties were mainly addressed by modifying the sample design to reduce the number of such schools selected rather than to exclude them (see Chapter 8 for further details on weighting). Schools where all students would be excluded through the within-school exclusion categories could be excluded up to a maximum of 2% as previously noted.

Within participating schools, all PISA-D-eligible students (i.e., born within the defined time period and in grades 7 or higher) were to be listed. From this, either a sample of TCS students was randomly selected or all students were selected if there were fewer than TCS students (as described in the "Student Sampling" section). The lists had to include students deemed to meet any of the categories for exclusion and a variable maintained to briefly describe the reason for exclusion, where applicable. This made it possible to estimate the size of the within-school exclusions from the sample data.

It was understood that the exact extent of within-school exclusions would not be known until the within-school sampling data were returned from participating schools and sampling weights computed. Participating country projections for within-school exclusions provided before school sampling were known to be estimates.

NPMs were made aware of the distinction between within-school exclusions and non-response. Students who could not take the PISA-D achievement tests because of a permanent condition were to be excluded and those with a temporary impairment at the time of testing, such as a broken arm, were treated as non-respondents along with other absent sampled students.

Exclusions by country are documented in Chapter 11.

Sampling frame

All NPMs were required to construct a school sampling frame to correspond to their national defined target population. The school sampling frame was defined by the *School Sampling Preparation Manual* as a frame that would provide complete coverage of the national defined target population without being contaminated by incorrect or duplicate entries or entries

referring to elements that were not part of the defined target population. It was expected that the school sampling frame would include any school with PISA-D eligible students, even those schools which might later be excluded or deemed ineligible because they had no PISA-D-eligible students at the time of data collection. The quality of the sampling frame directly affects the survey results through the schools' probabilities of selection and therefore their weights and the final survey estimates. NPMs were therefore advised to be diligent and thorough in constructing their school sampling frames.

All countries participating in PISA-D used school-level sampling frames as their first stage of sample selection. The *School Sampling Preparation Manual* indicated that the quality of sampling frames would largely depend on the accuracy of the approximate enrolment of 15-year-olds available (ENR) for each first-stage sampling unit. A suitable ENR value was a critical component of the sampling frames since selection probabilities were based on it. The best ENR for PISA-D was the number of currently enrolled 15-year-old students. Current enrolment data, however, were rarely available at the time of school sampling, which meant using alternatives. Most countries used the first-listed available option from the following list of alternatives:

- student enrolment in the target age category (15-year-olds) from the most recent year of data available;
- if 15-year-olds tended to be enrolled in two or more grades, and the proportions of students who were aged 15 in each grade were approximately known, the 15-year-old enrolment could be estimated by applying these proportions to the corresponding grade-level enrolments;
- the grade enrolment of the modal grade for 15-year-olds; and
- total student enrolment, divided by the number of grades in the school.

The *School Sampling Preparation Manual* noted that if reasonable estimates of ENR did not exist or if the available enrolment data were out of date, schools might have to be selected with equal probabilities which might require an increased school sample size. However, no countries needed to use this option.

Besides ENR values, NPMs were instructed that each school entry on the frame should include at minimum:

- school identification information, such as a unique numerical national identification, and contact information such as name, address, and phone number; and
- coded information about the school, such as region of country, school type, and extent of urbanisation, which would be used as stratification variables.

Stratification

Prior to sampling, schools were to be ordered, or stratified, in the sampling frame. Stratification consists of classifying schools into similar groups according to selected variables referred to as stratification variables. Stratification in PISA-D was used to:

- improve the efficiency of the sample design, thereby making the survey estimates more reliable;
- apply different sample designs, such as disproportionate sample allocations, to specific groups of schools, such as those in states, provinces, or other regions;
- ensure all parts of a population were included in the sample; and
- ensure adequate representation of specific groups of the target population in the sample.

There were two types of stratification utilised: explicit and implicit. Explicit stratification consists of grouping schools into strata that will be treated independently from one another, as if they were separate school sampling frames. Examples of explicit stratification variables could be states or regions of a country. Implicit stratification consists essentially of sorting the schools uniquely within each explicit stratum by a set of designated implicit stratification variables. Examples of implicit stratification variables could be type of school, urbanisation, or minority composition. Implicit stratification is a way of ensuring a strictly proportional sample allocation of schools across all the groups used for implicit stratification. It can also lead to improved reliability of survey estimates, provided that the implicit stratification variables being considered are correlated with PISA-D achievement at the school level. Guidelines on choosing stratification variables that would possibly improve the sampling were provided in the *FT Sampling Guidelines Manual.*[2]

Table 4.1 provides the explicit stratification variables used by each country, as well as the number of explicit strata found within each country. For example, Cambodia had two explicit strata using urbanicity (known as locations), which were then further delineated by three school types and also had one explicit stratum for certainty selections, so that there were nine explicit strata in total. Variables used for implicit stratification and the respective number of levels can also be found in Table 4.1.

As the sampling frame was always finally sorted by school size, school size was also an implicit stratification variable, though it is not listed in Table 4.1. The use of school size as an implicit stratification variable provides a degree of control over the student sample size so as to possibly avoid the sampling of too many relatively large schools or too many relatively small schools.

Table 4.1 **Stratification variables used in PISA-D**

Country	Explicit stratification variables	Number of explicit strata	Implicit stratification variables
Cambodia	Location (2); School types (3); School zones (4)	25	School Management (2); Shifts (2)
Ecuador	Urbanicity (2); School Management (4)	8	Province (25); Academic calendar (2); ISCED levels (3)
Guatemala	Área (2); Sector (4)	9	ISCED (3); Modality (4)

Honduras	Urbanicity (2); School Management (2); National School Classification (2)	9	Gender (5); Department (18)
Paraguay	School sector (3); School area (2); School size (3)	18	Region (5)
Senegal	Academic (16); Urbanicity (2);	30	School Management (2); Inspection (59); Gender (3)
Zambia	Province (10); Location (2)	20	School Type (4)

Assigning a measure of size to each school

for the probability proportional to size sampling method used for PISA-D, a Measure of Size (MOS) derived from ENR was established for each school on the sampling frame. MOS was generally constructed as: MOS = max(ENR, TCS). This differed slightly in the case of small schools treatment, which is discussed later. Thus, the measure of size was equal to the enrolment estimate (ENR), unless enrolment was less than the TCS, in which case the measure of size was set equal to the target cluster size.

As schools were sampled with probability proportional to size, setting the measure of size of small schools to 42 students was equivalent to drawing a simple random sample of small schools. That is, small schools would have an equally likely chance of being selected to participate. However, the number of small schools in the sample has an impact on important operational and budgetary matters for each country. Therefore, some further adjustments were made after taking these factors into consideration. Please see "Treatment of Small Schools" for details on how small schools were sampled.

School sample selection

School sample allocation over explicit strata

The total number of schools to be sampled in each country needed to be allocated among the explicit strata so that the expected proportion of students in the sample from each explicit stratum was approximately the same as the population proportions of PISA-D-eligible students in each corresponding explicit stratum. There were two exceptions. If very small schools required undersampling, students in them had smaller percentages in the sample than in the population. To compensate for the resulting loss of sample, the large schools had slightly higher percentages in the sample than the corresponding population percentages. The other exception occurred if only one school was allocated to any explicit stratum. In this case, two schools were allocated for selection in the stratum to aid with variance estimation.

Sorting the sampling frame

The *School Sampling Preparation Manual* indicated that, prior to selecting the school sample, schools in each explicit stratum were to be sorted by a limited number of variables chosen for implicit stratification, and finally by the ENR value within each implicit stratum. The schools were first to be sorted by the first implicit stratification variable, then by the second implicit

stratification variable within the levels of the first implicit stratification variable, and so on, until all implicit stratification variables were used. This gave a cross-classification structure of cells, where each cell represented one implicit stratum on the school sampling frame. The sort order was alternated between implicit strata, from high to low and then low to high, and so on, through all implicit strata within an explicit stratum.

Determining which schools to sample

The PPS-systematic sampling method used in PISA-D first required the computation of a sampling interval for each explicit stratum. This calculation involved the following steps:

- recording the total measure of size, S, for all schools in the sampling frame for each specified explicit stratum;
- recording the number of schools, D, to be sampled from the specified explicit stratum, which was the number allocated to the explicit stratum;
- calculating the sampling interval, I, as follows: $I = S/D$;
- including in the sample all schools for which the school's size measure exceed I (known as certainty schools)*;*
- removing certainty schools from the frame, recalculating S, D, and I; and
- recording the sampling interval, I, to four decimal places.

Next, a random number had to be generated for each explicit stratum**.** The generated random number (RN) was from a uniform distribution between zero and one and was to be recorded to four decimal places.

The next step in the PPS selection method in each explicit stratum was to calculate selection numbers - one for each of the D schools to be selected in the explicit stratum. Selection numbers were obtained using the following method:

- Obtaining the first selection number by multiplying the sampling interval, I, by the random number, *RN*. This RN number is a random number between zero and one, and to 4 decimal places. This first selection number was used to identify the first sampled school in the specified explicit stratum.
- Obtaining the second selection number by adding the sampling interval, I, to the first selection number. The second selection number was used to identify the second sampled school.
- Continuing to add the sampling interval, I, to the previous selection number to obtain the next selection number. This was done until all specified line numbers (1 through D) had been assigned a selection number.

Thus, the first selection number in an explicit stratum was $RN \times I$, the second selection number was $(RN \times I) + I$, the third selection number was $(RN \times I) + I + I$, and so on.

Selection numbers were generated independently for each explicit stratum, with a new random

number generated for each explicit stratum.

Identifying the sampled schools

The next task was to compile a cumulative measure of size in each explicit stratum of the school sampling frame that assisted in determining which schools were to be sampled. Sampled schools were identified as follows.

Let Z denote the first selection number for a particular explicit stratum. It was necessary to find the first school in the sampling frame where the cumulative *MOS* equalled or exceeded Z. This was the first sampled school. In other words, if C_s was the cumulative *MOS* of a particular school S in the sampling frame and $C_{(s-1)}$ was the cumulative *MOS* of the school immediately preceding it, then the school in question was selected if: C_s was greater than or equal to Z, and $C_{(s-1)}$ was strictly less than Z. Applying this rule to all selection numbers for a given explicit stratum generated the original sample of schools for that stratum.

<div style="border: 1px solid black; padding: 10px;">

<div align="center">Box 4.1 **Illustration of probability proportional to size (PPS) sampling**</div>

To illustrate these steps, suppose that in an explicit stratum in a participant country, the PISA-D-eligible student population is 105,000, then:

- the total measure of size, *S*, for all schools is 105,000;
- the number of schools, *D*, to be sampled is 150;
- calculating the sampling interval, *I*, 105,000/150 = 700;
- generate a random number, *RN*, 0.3230;
- the first selection number is 700 X 0.3230 = 226.1. This first selection number is used to identify the first sampled school in the specified explicit stratum; and
- the second selection number is 226 + 700 = 926.1. The second selection number was used to identify the second sampled school.

The third selection number is 926 + 700 = 1,626.1. The third selection number was used to identify the third sampled school, and so on until the end of the school list is reached. This will result in a school sample size of 150 schools.

The table below also provides these example data. The school that contains the generated selection number within its cumulative enrolment is selected for participation.

School	*MOS*	Cumulative *MOS* (C_s)	Selection Number	
001	550	550	226.1	Selected
002	364	914		
003	60	974	926.1	Selected
004	93	1067		
005	88	1155		
006	200	1355		
007	750	2105	1626.1	Selected
008	72	2177		
009	107	2284		
010	342	2626	2326.1	Selected
011	144	2770		
⋮	⋮	⋮	⋮	⋮

</div>

Identifying replacement schools

Each sampled school in the Main Survey was assigned two replacement schools from the school sampling frame, if possible, identified as follows. For each sampled school, the schools immediately preceding and following it in the explicit stratum, which was ordered within by the implicit stratification, were designated as its replacement schools. The school immediately following the sampled school was designated as the first replacement and labelled R_1, while the school immediately preceding the sampled school was designated as the second replacement and labelled R_2. Exceptions were allowed if a sampled school happened to be the last school listed in an explicit stratum. In this case the two schools immediately preceding it were designated as replacement schools. Similarly, for the first school listed in an explicit stratum, the two schools immediately following it were designated as replacement schools.

In certain situations, there could potentially be problems when trying to identify two replacement schools for each sampled school. This could happen when some very large sampled schools appeared close to each other in the sampling frame, or when it was only possible to assign a single replacement school, or even none, because two consecutive schools in the sampling frame were sampled. In such cases, the *School Sampling Preparation Manual,* stated that a replacement school could be allowed to be the potential replacement for two sampled schools. However, due to the general expectation of very good school response in all countries, and to avoid operational complexity, sample selection was completed with fewer than two assigned replacements for schools in this situation. Countries were alerted to any occurrences of this and were asked to confirm that this was acceptable before the sample selection was approved.

Assigning school identifiers

To keep track of sampled and replacement schools in the PISA-D database, each was assigned a unique three-digit school code sequentially numbered starting with one within each explicit stratum (each explicit strata was numbered with a separate two-digit stratum code). For example, if 150 schools were sampled from a single explicit stratum, they were assigned identifiers from 001 to 150. First replacement schools in the Main Survey were assigned the school identifier of their corresponding sampled schools, with increments of 300. For example, the first replacement school for sampled school 023 was assigned school identifier 323. Second replacement schools in the Main Survey were assigned the school identifier of their corresponding sampled schools, but with an increment of 600. For example, the second replacement school for sampled school 136 took the school identifier 736.

Tracking sampled schools

NPMs were encouraged to make every effort to confirm the participation of as many sampled schools as possible to minimise the potential for non-response biases. Each sampled school that did not participate was replaced if possible.

Treatment of small schools

In PISA-D, schools were classified as very small, moderately small, or large. A school was classified as *large* if it had an ENR at or above the TCS (42 students in most countries). A *moderately small* school had an ENR in the range of one-half the TCS to TCS (21 to less than 42 students in most

countries). A *very small* school had an ENR less than one-half the TCS (20 students or fewer in most countries)**.** These schools were further classified as either *P1 very small schools* with an ENR of zero, one, or two students, or *P2 very small schools* with an ENR greater than two students but less than one-half the TCS. Unless they received special treatment in the sampling, the occurrence of small schools in the sample would reduce the sample size of students for the national sample to below the desired target because the within-school sample size would fall short of expectations. A sample with many small schools could also be an administrative burden with many testing sessions with few students. To minimise these problems, procedures were devised for managing small schools in the sampling frame.

To balance the two objectives of selecting an adequate sample of small schools but not too many small schools so as to hurt student yield, a procedure was recommended that assumed the underlying idea of undersampling the *P2 very small* schools by a factor of two and undersampling the *P1 very small* schools by a factor of four and to proportionally increase the number of large schools to sample. Additionally, in some countries, *P2 very small* schools were also undersampled by a factor of four to keep the total number of sampled schools manageable, after the proportional increase of the number of large schools to sample.

To determine whether very small schools should be undersampled and if the sample size needed to be increased to compensate for small schools, the following test was applied.

- If the percentage of students in very small schools was 1 percent or *more*, then very small schools were undersampled and the school sample size increased, sufficient to maintain the required overall yield.
- If the percentage of students in very small schools was *less* than 1 percent, and the percentage of students in moderately small schools was 4 percent or *more*, then there was no required undersampling of very small schools, but the school sample size was increased sufficiently to maintain the required overall yield.

If none of these conditions were true, then the small schools contained such a small proportion of the PISA-D population that they were unlikely to reduce the sample below the desired target. In this case, no undersampling of very small schools was needed nor an increase to the school sample size to compensate for small schools.

The approach included the condition where the percentage of *schools* on the frame that were the P1 schools was 20 percent or more. This approach was for the infrequent situation where very small schools (ENR < TCS/2) overall contain less than 1 percent of total frame enrolment while at the same time these very smallest schools accounted for a large percentage of total schools on the frame. If this condition was met and no undersampling was otherwise required based on the percentage of enrolment in very small schools, very small schools were undersampled to avoid having too many of them in the school sample. Even though undersampling can reduce the number of them in the sample from what could be expected without undersampling, when very small schools account for such a large percentage of schools on the frame, it is likely that a relatively large number of them (but not a large proportion) will be selected. A minor increase to

the sample size was needed in this case to safeguard the needed student sample size.

Since the number of very small schools was to be controlled in the sample without creating explicit strata for these small schools, this was accomplished by assigning an MOS- of TCS/2 to the *P2 very small* schools and a measure of size equal to the TCS/4 for the *P1 very small* schools. For the few countries where additional undersampling was done to result in a manageable total number of sampled schools, the measure of size for the *P2 very small* schools was also set to TCS/4. In effect, very small schools with a measure of size equal to TCS/2 were undersampled by a factor of two (school probability of selection reduced by half), and the very small schools with a measure of size equal to TCS/4 were undersampled by a factor of four (school probability of selection reduced by three-fourths). This was accomplished as follows and was a standard procedure followed in all countries.

The formulae below assumed an initial target school sample size of 150 and a target student sample size of 6,300.

- Step 1: From the complete sampling frame, find the proportions of total ENR that come from *P1 very small* schools, *P2 very small* schools), moderately small schools (*Q*), and large schools (*R*). Thus, $P1+P2+Q+R=1$
- Step 2: Calculate the value *L*, where $L= 1.0 + 3(P1)/4 + (P2)/2$. Thus *L* is a positive number slightly more than 1.0.
- Step 3: The minimum sample size for large schools is equal to $150 \times R \times L$, rounded up to the nearest integer. It may need to be enlarged because of national considerations, such as the need to achieve minimum sample sizes for geographic regions or certain school types.
- Step 4: Calculate the mean value of ENR for moderately small schools (*MENR*), and for very small schools (*V1ENR* and *V2ENR*). *MENR* is a number in the range of TCS/2 to TCS, *V2ENR* is a number larger than two but no greater than TCS/2, and *V1ENR* is a number in the range of zero to two.
- Step 5: The number of schools that must be sampled from the moderately small schools is given by: $(6,300 \times Q \times L)/(MENR)$.
- Step 6: The number of schools that must be sampled from the very small schools (type *P2*) is given by: $(3,150 \times P2 \times L)/(V2ENR)$.
- Step 7: The number of schools that must be sampled from the very small schools (type *P1*) is given by: $(1,575 \times P1 \times L)/(V1ENR)$.

To illustrate the steps, suppose that in a participant country, the TCS is equal to 42 students, with 10% of the total enrolment of 15-year-olds in moderately small schools, and 5% in each type of very small schools, *P1* and *P2*. Suppose that the average enrolment in moderately small schools is 25 students, in very small schools (type *P2*) it is 12 students, and in very small schools (type *P1*) it is 1.5 students.

- Step 1: The proportions of total ENR from very small schools is *P1* = 0.05 and *P2* = 0.05,

from moderately small schools is Q = 0.1, and from large schools is R = 0.8. The proportion of the very smallest schools on the frame was not more than 20%. It can be shown that 0.05 + 0.05 + 0.1 + 0.8 = 1.0.

- Step 2: Calculate the value L. L= 1.0 + 3(0.05)/4 + (0.05/2). Thus L = 1.0625.

- Step 3: The minimum sample size for large schools is equal to 150 × 0.8 × 1.0625 = 127.5. That is, at least 128 (rounded up to the nearest integer) of the large schools must be sampled.

- Step 4: The mean value of ENR for moderately small schools ($MENR$) is given in this example as 25, very small schools of type $P2$ ($V2ENR$) as 12, and very small schools of type $P1$ ($V1ENR$) as 1.5.

- Step 5: The number of schools that must be sampled from the moderately small schools is given by (6,300 × 0.1 × 1.0625)/25 = 26.8. At least 27 (rounded up to the nearest integer) moderately small schools must be sampled.

- Step 6: The number of schools that must be sampled from the very small schools (type $P2$) is given by (3,150 × 0.05 × 1.0625)/12 = 13.9. At least 14 (rounded up to the nearest integer) very small schools of type $P2$ must be sampled.

- Step 7: The number of schools that must be sampled from the very small schools (type $P1$) is given by (1,575 × 0.05 × 1.0625)/1.5 = 55.8. At least 56 (rounded up to the nearest integer) very small schools of type $P1$ must be sampled.

Combining these different sized school samples gives a total sample size of 128 + 27 + 14 + 56 = 225 schools. Before considering school and student non-response, the larger schools will yield an initial sample of approximately 128 × 42 = 5,376 students. The moderately small schools will give an initial sample of approximately 27 × 25 = 675 students, very small schools of type $P2$ will give an initial sample size of approximately 14 × 12 = 168 students, and very small schools of type $P1$ will give an initial sample size of approximately 56 × 1.5 = 84 students. The total expected sample size of students is therefore 5,376 + 675 + 168 + 84 = 6,303. This is a representative example for a PISA-D country assuming a student response rate of 85% to attain a minimum of 5,250 completed assessments.

This procedure, called small school analysis, was done not just for the entire school sampling frame, but for each individual explicit stratum. An initial allocation of schools to explicit strata provided the starting number of schools and students to project for sampling in each explicit stratum. The small school analysis for a single unique explicit stratum indicated how many very small schools of each type (assuming undersampling, if needed), moderately small schools and large schools would be sampled in that stratum. Together, these provided the final sample size, n, of schools to select in the stratum. Based on the stratum sampling interval and random start, large, moderately small, and very small schools were sampled in the stratum to a total of n sampled schools. Because of the random start, it was possible to have more or less than expected of the very small schools of either type, $P1$ or $P2$, of the moderately small schools, and of the large schools. The total number of sampled schools however was fixed at n, and the number of expected students to be sampled was always approximate to what had been projected from the unique stratum small school analysis

School sampling

PISA-D Technical Standard 1.13 states that the international contractor should select the school samples unless otherwise agreed upon (see Annex D). No countries selected their own school sample; therefore all participating countries school samples were selected by and checked in detail by the international contractor. To enable this, all countries were required to submit sampling information on forms associated with the following various sampling tasks:

- language(s) of instruction were identified in Sampling Task 0 at the time of the Field Trial, including the determination of which language(s) would be used for the Field Trial and Main Survey assessments;

- time of testing and age definition for both the Field Trial and Main Survey were captured on Sampling Task 1 (see below) at the time of the Field Trial, with updates being possible before the Main Survey;

- information about stratification for the Field Trial and Main Survey was recorded on Sampling Task 2;

- forms or data associated with Sampling Tasks 3, 4, 5, and 6 were all for the Field Trial;

- the national desired target population information for the Main Survey was captured on the form associated with Sampling Task 7a;

- information about the defined national target population (including expected school-level and within-school exclusions) was recorded on the form associated with Sampling Task 7b;

- the description of the sampling frame was noted on the form associated with Sampling Task 8a; and

- the school sampling frame was created in one spreadsheet and the list of any excluded schools in a second spreadsheet associated with Sampling Task 8b.

The international contractor completed school sampling and, along with the school sample, returned other information (small school analyses, school allocation, and a spreadsheet that countries could use for tracking school participation). Table 4.2 provides a summary of the information required for each sampling task and the timetables (which depended on national assessment periods).

Table 4.2 **Schedule of school sampling activities**

Activity	Submit to Consortium	Due Date
Update time of testing and age definition of population to be tested	Sampling Task 1 – time of testing and age definition	Update what was submitted at the time of the Field Trial (FT), two months before the school sample is to be selected
Finalise explicit and implicit stratification variables	Sampling Task 2 – stratification and other information	Update what was submitted at the time of the FT, two months before the school sample is to be selected
Define national desired target population	Sampling Task 7a – national desired target population	Submit two months before the school sample is to be selected
Define national defined target population; list school-level exclusions	Sampling Task 7b – national defined target population	Submit two months before the school sample is to be selected
Create and describe sampling frame	Sampling Task 8a – sampling frame description	Submit two months before the school sample is to be selected
Submit sampling frame	Sampling Task 8b – sampling frame (in one Excel sheet), and excluded schools (in another Excel sheet)	Submit two months before the school sample is to be selected
Decide how to treat small schools	Treatment of small schools	The international contractor completes and returns this information to the NPM about one month before the school sample is to be selected
Finalise sample size requirements	Sampling Task 9 – sample allocation by explicit strata	The international contractor completes and returns this information to the NPM about one month before the school sample is to be selected
Describe population within strata	Population counts by strata	The international contractor completes and returns this information to the NPM when the school sample is sent to the NPM
Select the school sample	Sampling Task 10 – school sample selection	The international contractor returns the sampling frame to the NPM with sampled schools and their replacement schools identified and with PISA IDs assigned when the school sample is selected
Review and agree to the sampling form required as input to *MyMWestra*	Sampling Task 11 – reviewing and agreeing to the Sampling Form for *MyMWestra* (SFMW)	Countries had one month after their sample was selected to agree to their SFMW
Submit sampling data	Sampling Task 12 – school participation information and data validity checks	Submit in *MyMWestra* within one month of the end of the data collection period

Once received from each participating country, each set of information was reviewed and feedback was provided to the country. Forms were only approved after all criteria were met. Approval of deviations was only given after discussion and agreement by the international contractors. In cases where approval could not be granted, countries were asked to make revisions to their sample design and sampling forms and resubmit.

Checks that were performed when monitoring each sampling task follow. Although all sampling tasks were checked in their entirety, the below paragraphs contain matters that were explicitly examined. Sampling Tasks 3–6 were only for the Field Trial and are not included.

Sampling Task 0: Languages of instruction

- Countries were asked to confirm that all possible languages of instruction were documented.
- The existence of international/foreign schools was asked about.
- Checks were done on the appropriate inclusion of languages for the Field Trial and Main Survey.
- Languages that were planned for Main Survey exclusion were scrutinised.

Sampling Task 1: Time of testing and age definition

- Assessment dates had to be appropriate for the selected target population dates.
- Assessment dates could not cover more than a 42-day period unless agreed upon.
- Assessment dates could not be within the first six weeks of the academic year.
- Any issues with the timing of the academic calendar, with respect to the PISA-D field period, were discussed.
- If assessment end dates were close to the end of the target population birth-date period, NPMs were alerted not to conduct any make-up sessions beyond the date when the population birth dates were valid.

Sampling Task 2: Stratification (and other information)

- Each participating country used explicit strata to group similar schools together to reduce sampling variance and to ensure representativeness of students in various school types using variables that might be related to outcomes. The international contractor assessed each country's choice of explicit stratification variables. If a country was known to have school tracking or distinct school programmes and they were not among the explicit stratification variables, a suggestion was made to include this type of variable.
- Levels of variables and their codes were checked for completeness.
- If no implicit stratification variables were noted, suggestions were made about ones that might be used. In particular, if a country had single gender schools and school gender was not among the implicit stratification variables, a suggestion was made to include this

type of variable to ensure no sample gender imbalances. Similarly, if there were International Standard Classification of Education (ISCED) school level splits, the ISCED school level was also suggested as an explicit or implicit stratification variable.

Sampling Task 7a: National desired target population

- The total national number of 15-year-olds of participating countries was compared with preliminary information provided by the country at the beginning of the PISA-D project. Differences were queried.

- Large deviations between the total national number of 15-year-olds and the enrolled number of 15-year-olds were questioned.

- Any population to be omitted from the international desired population was noted and discussed.

- Data sources and the year of the data were required. If websites were provided with an English page option, the submitted data was verified against those sources.

Sampling Task 7b: National defined target population

- Reasons for excluding schools for reasons other than special education needs were thoroughly discussed and checked for appropriateness (i.e., some operational difficulty in assessing the school). In particular, school-level language exclusions were closely examined to check correspondence with what had been noted about language exclusions on Sampling Task 0.

- The number and percentage of students to be excluded at the school level and whether the percentage was less than the guideline for maximum percentage allowed for such exclusions were checked.

- Reasonableness of assumptions about within-school exclusions was assessed. If there was an estimate noted for "other", the country was queried regarding reasonableness about what the "other" category represented. If it was known the country had schools where some of the students received instruction in minority languages not being tested, an estimate for the within-school exclusion category for "no materials available in the student's language of instruction" was necessary.

- Form calculations were verified through built-in data checks, and the overall coverage figures were assessed.

- If it was noted that there was a desire to exclude schools with only one or two PISA-D-eligible students at the time of contact, then the school sampling frame was checked for the percentage of population that would be excluded. If countries had met the 2.5% school-exclusion guideline, and if these schools would account for not more than 0.5% and if within-school exclusions were within 2.5%, then the exclusion of these schools at the time of contact was agreed upon with the understanding that such exclusion not cause entire strata to be missing from the student data.

- The population figures on this form after school-level exclusions were compared against the aggregated school sampling frame enrolment. School-level exclusion totals also were

compared to those tabulated from the excluded school sheet of the sampling frame: ST8B. Differences were queried.

- Data sources and the year of the data were required. If websites were provided with an English page option, the submitted data was verified against those sources.

Sampling Task 8a: Sampling frame description

- The type of school-level enrolment estimate and the year of data availability were discussed thoroughly, to ensure the best possible source was used for each country.

- Countries were asked to provide information for each of various school types,[2] whether those schools were included on or excluded from the sampling frame, or the country did not have any of such schools. The information was matched to the different types of schools containing PISA-D students noted on Sampling Task 2. Any discrepancies were queried.

- Any school types noted as being excluded were verified as school-level exclusions on the Sampling Task 7b form. Any discrepancies were queried.

Sampling Task 8b: Sampling frame

- On the spreadsheet for school-level exclusions, the number of schools and the total enrolment figures, as well as the reasons for exclusion, were checked to ensure correspondence with values reported on the Sampling Task 7b form detailing school-level exclusions. It was verified that this list of excluded schools did not have any schools that were excluded for having only one or two PISA-D-eligible students, as these schools were not to be excluded from the school sampling frame. Checks were done to ensure that excluded schools did not still appear on the other spreadsheet containing the school sampling frame.

- All units on the school sampling frame were confirmed to be those reported on the Sampling Task 2 as sampling frame units.

- NPMs were queried about whether they had included schools with grades 7 or 8, or in some cases those with grades 10 or higher, which could potentially have PISA-D-eligible students at the time of assessment even if the school currently did not have any.

- NPMs were queried about whether they had included vocational or apprenticeship schools, schools with only part-time students, international or foreign schools, schools not under the control of the Ministry of Education, or any other irregular schools that could contain PISA-D-eligible students at the time of the assessment, even if such schools were not usually included in other national surveys.

- The frame was checked for all required variables: a national school identifier with no duplicate values, a variable containing the school enrolment of PISA-D-eligible students, and all the explicit and implicit stratification variables. Stratification variables were

---

[2]   These include schools with multiple languages of instruction, vocational schools, technical schools, agriculture schools, and schools with only part-time students, schools with multiple shifts, and so on.

checked to make sure none had missing values and only had levels as noted on Sampling Task 2.

- Any additional school sampling frame variables were assessed for usefulness. In some instances, other variables were noted on the school frame that might also have been useful for stratification.
- The frame was checked for schools with only one or two PISA-D-eligible students to ensure these were not erroneously excluded from the sampling frame.
- The frame was checked for schools with zero enrolment. If there were none, this was assessed for reasonableness. If some existed, it was verified with the NPM that these schools could possibly have PISA-D-eligible students at the time of the assessment.

Sampling Task 9: Treatment of small schools and the sample allocation by explicit strata

- All explicit strata had to be accounted for on the form for Sampling Task 9.
- All explicit strata population entries were compared to those determined from the sampling frame.
- All small school analysis calculations were verified.
- Country-specified sample sizes were monitored and revised if necessary to be sure minimum sample sizes were being met.
- The calculations for school allocation were checked to ensure that schools were allocated to explicit strata based on explicit stratum student percentages and not explicit stratum school percentages, that all explicit strata had at least two allocated schools, and that no explicit stratum had only one remaining non-sampled school.
- It was verified that the allocation matched the results of the explicit strata small school analyses, with allowances for random deviations in the numbers of very small, moderately small, and large schools to be sampled in each explicit stratum.
- The percentage of students in the sample for each explicit stratum had to be approximate to the percentage in the population for each stratum.
- The overall number of schools to be sampled was checked to ensure that at least 150 schools would be sampled.
- The overall number of students to be sampled was checked to ensure that at least 5,250 students would be sampled.

Sampling Task 10: School sample selection

- All calculations were verified.
- Particular attention was paid to the required four decimal places for the sampling interval and the generated random number.
- The frame was checked for proper sorting according to the implicit stratification scheme, for enrolment values, and the proper assignment of the measure of size value, especially for very small and moderately small schools. The assignment of replacement schools and PISA-D identification numbers were checked to ensure that all rules established in the

Main Survey School Sampling Preparation Manual were adhered to.

Sampling Task 11: Reviewing and agreeing to the Sampling Form

- The form for Sampling Task 11 was prepared as part of the sample selection process. After the international contractor verified that all entries were correct, NPMs had to perform the same checks and agree to the content in this form.

Sampling Task 12: School participation and data validity checks

- Extensive checks were completed on Sampling Task 12 data since it would inform the weighting process. Checks were done to ensure that school participation statuses were valid, student participation statuses had been correctly assigned, and all student sampling data required for weighting were available and correct for all student sampling options. Quality checks also highlighted schools having only one grade with PISA-D-eligible students, only one gender of PISA-D-eligible students, or schools that had noticeable differences in enrolled student counts than expected based on sampling frame enrolment information. Such situations were queried.
- Uneven distributions of student birth months were queried.

Student samples

Student selection procedures in the Main Survey were the same as those used in the Field Trial. Student sampling was undertaken by the international contractor, using lists of all PISA-D-eligible students in each school that had agreed to participate, which were submitted to the International Contractor through the MyMWestra system. These lists could have been prepared at national, regional, or local levels as data files, computer-generated listings, or by hand, depending on who had the most accurate information. Since it was important that the student sample be selected from accurate, complete lists, the lists needed to be prepared slightly in advance of the testing period and had to list all PISA-D-eligible students. It was suggested that the lists be received one to two months before the testing period so that the international contractor would have adequate time to select the student samples.

Preparing a list of age-eligible students

Each school participating in PISA-D had to prepare a list of age-eligible students that included all 15-year-olds (using the appropriate 12-month age span agreed upon for each participating country) in international grades 7 or higher. The following were considered important:

- Age-eligible students were all students born in the appropriate 12-month age span agreed upon for the participating country.
- The list was to include students who might not be tested due to a disability or limited language proficiency.
- Students who could not be tested were to be excluded from the assessment after the student listing form was created and after the student sample was selected. It was stressed to National Centres that students were to be excluded after the student sample

was drawn, not prior.

- It was suggested that schools retain a copy of the student list in case the NPM had to contact the school with questions.
- Student lists were to be up to date, as close as possible to the time of student sampling rather than a list prepared at the beginning of the school year.

Selecting the student sample

Once NPMs received the list of PISA-D-eligible students from all schools and submitted to the international contractor, all the student samples were selected and the list of selected students returned to the National Centre via Student Tracking Forms. The NPM would then send the tracking forms to each school. An equal probability sample of PISA-D students was selected, using systematic sampling, where the lists of students was first sorted by grade and gender. For PISA-D, all countries were required to use MyMWestra, the International Contractor software, for all student sampling related activities and the contractor selected all student samples.

Preparing instructions for excluding students

PISA-D was a timed assessment administered in the instructional language of each participating country and designed to be as inclusive as possible. In the interest of ensuring that PISA-D was as inclusive as possible, student participation and reasons for exclusion were separately coded in the student tracking form. This allowed for students classified as SEN to be included when their SEN was not severe enough to be a barrier to their participation. The participation status could therefore detail, for example, that a student participated and was not excluded for SEN reasons even though the student was noted with a special educational need. Any student whose participation status indicated they were excluded for SEN reasons had to have a SEN code that explained the reason for exclusion. It was important that these criteria were followed strictly for the study to be comparable within and across participating countries. School co-ordinators and test administrators were told to include students when in doubt.

School co-ordinators were asked to review the Student Tracking Forms and indicate which students, if any, had a SEN classification. The codes for this classification are displayed in Table 4.3.

Table 4.3 **Codes for special educational needs**

Description	Code
Functional disability—student has a moderate to severe permanent physical disability	**1**
Cognitive, behavioural, or emotional disability—in the opinion of qualified staff, student has a cognitive, behavioural, or emotional disability	**2**
Limited assessment language experience—student is not a native speaker of any of the languages of the assessment in the country and has limited proficiency in these languages	**3**
<Optional additional category, defined by NPM and approved by contractors>	**<4>**
No materials available in language of instruction—There are no materials available in the language in which the student is taught	**5**

Once SEN students were identified, clear exclusion and inclusion guidelines were provided to the school co-ordinators. It was expected that most SEN students would participate in PISA-D. These guidelines for exclusion/inclusion are presented in Table 4.4.

Table 4.4 **Guidelines—Exclusion/inclusion of students with special educational needs**

Group	Students who should be INCLUDED	Students who may be EXCLUDED
Functional (physical or sensory) disability	The student can respond to the assessment. *Record Code 1 in column "SEN."* *Leave column "N/P PISA" blank.*	The student has a moderate to severe permanent physical disability such that **he or she cannot participate in the PISA testing situation**. *Record Code 1 in column "SEN."* *Record Code 3 in column "N/P PISA."*
General learning disability, behavioural, or emotional disability	The student can respond to the assessment. He or she should **NOT** be excluded solely because of poor academic performance or disciplinary problems. *Record Code 2 in column "SEN."* *Leave column "N/P PISA" blank.*	The student has a cognitive, behavioural, or emotional disability such that in the opinion of qualified staff, **he or she cannot participate in the PISA testing situation**. This includes students who are cognitively, behaviourally, or emotionally unable to follow even the general instructions of the assessment. *Record Code 2 in column "SEN."* *Record Code 3 in column "N/P PISA."*
Insufficient assessment language experience	The student does not meet **ALL** three criteria in the **Students who may be EXCLUDED** column . *Record Code 3 in column "SEN."* *Leave column "N/P PISA" blank.*	The student meets **ALL** of the following three criteria: - not being a native speaker in the assessment language, - having limited proficiency in the assessment language, and - having received **less than 1 year of instruction** in the assessment language. *Record Code 3 in column "SEN."* *Record Code 3 in column "N/P PISA."*
Assessment is not available in student's language of instruction	The student does not receive the majority of instruciton in a language other than <language(s) of assessment>. *Leave column "N/P PISA" blank.*	The student receives the majority of instruction in a language other than <language(s) of assessment>. *Code 3 in column "N/P PISA."*

PISA-D NPMs adapted the codes and guidelines. Only minor adaptions to the guidelines were allowed. The National Centre also determined if it wanted to use Code 4, the optional SEN category. This required contractor approval. No PISA-D countries used this option. In some countries, Code 5 (no materials available in language of instruction) was not necessary.

The national operational definitions for within-school exclusions were to be clearly documented and submitted to the international contractor for review before testing.

Sending the Student Tracking Form to the School Co-ordinator and Test Administrator

The school co-ordinator needed to know which students were sampled in order to notify students, parents, and teachers, and in order to update information and to identify students to be excluded. It was recommended that the National Centre send Student Tracking Forms to test administrators and to school co-ordinators at least two weeks before testing began in order for the school to have adequate time to prepare for the assessment. (See Chapter 6).

Teacher samples

New for PISA 2015, a limited number of countries elected to take an international option in which teachers were sampled in each sampled school. Data from the teacher questionnaire (TQ) was intended to be used to add context to student data from the same school, that is, to describe the learning environment of typical 15-year-old students in the country. Therefore, the TQ focused on that grade level that most 15-year-old students in in the country attend, a.k.a. the national modal grade for 15-year-old students. If an adjacent grade level was attended by one third or more of 15-year-old students in the country, both grade levels were used as modal grades.

A teacher was defined as "one whose primary or major activity in the school is student instruction, involving the delivery of lessons to students. Teachers could work with students as a whole class in a classroom, in small groups in a resource room or one-to-one inside or outside regular classrooms."

In order to cover a broader variety of perspectives, and guarantee samples that were large enough, teachers who *could* or *would* be teaching the PISA modal grade in a later year were also considered to belong to the teacher target population. This applied also for teachers who had been teaching the modal grade in the past who were still in the school. Thus, sampling for teachers included *all* teachers that were *eligible* for teaching the modal grade—whether they were doing so currently, had done so before, or will/could do so in the future.

Teachers were listed and sampled in KeyQuest as either part of Population 4 (science teachers) or Population 5 (non-science teachers). The distinction between Population 4 and Population 5 was determined by the meaning of school science. School science included all school science courses referring to the domains of physics, chemistry, biology, earth science or geology, space science or astronomy, applied sciences, and technology, either taught in the curriculum as separate science subjects or taught within a single "integrated-science" subject. It did *not* include related subjects such as mathematics, psychology, economics, nor possible earth science topics included in geography courses.

Ten science teachers were sampled in schools having at least that many listed, or all, if there were not 10. Fifteen non-science teachers were sampled in schools having at least that many listed, or all if there were not 15. Within each teacher population (science and non-science), an equal probability sample of teachers was selected, using systematic sampling where the lists of teachers were first sorted by grade and gender, where grade had codes indicating whether the teacher was currently teaching the modal grade.

Definition of school

Although the definition of a "school" is difficult, PISA-D generally aimed to sample whole schools as the first stage units of selection, rather than programmes or tracks or shifts within schools, so that the meaning of "between school variance" would be more comparable across countries.

There were exceptions to this, such as when school shifts were actually more like separate schools than part of the same overall school. However, in some countries with school shifts this was not the case, and therefore whole schools were used as the primary sampling unit. Similarly, many countries had schools with different tracks/programs, but generally it was recommended again that the school as a whole should be used as the primary sampling unit.

Sampling units to be used on school-level frames were discussed with each country before the Field Trial. All countries used "whole school" as the unit of sampling.