4

Sample design

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

TARGET POPULATION AND OVERVIEW OF THE SAMPLING DESIGN

The desired base PISA target population in each country consisted of 15-year-old students attending educational institutions located within the country, in grades 7 and higher. This meant that countries were to include (i) 15-year-olds enrolled full-time in educational institutions, (ii) 15-year-olds enrolled in educational institutions who attended on only a part-time basis, (iii) students in vocational training types of programmes, or any other related type of educational programmes, and (iv) 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 persons schooled in the home, workplace or out of the country would occur and therefore these students 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 March 2006 and 31 August 2006, unless otherwise agreed, during which they would administer the assessment.

Further, testing was not permitted during the first six weeks of the school year because of a concern that student performance levels may be lower at the beginning of the academic year than at the end of the previous academic year, even after controlling for age.

The 15-year-old international target population was slightly adapted to better fit the age structure of most of the Northern Hemisphere countries. As the majority of the testing was 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 meant that in all countries testing in April 2006, the target population could have been defined as all students born in 1990 who were attending a school or other educational institution.

Further, a variation of up to one month in this age definition was permitted. This was done to allow a country testing in March or in May to still define the national target population as all students born in 1990. If the testing was to take place at another time until the end of August, the birth date definition had to be adjusted.

In all but one country, the sampling design used for the PISA assessment was a two-stage stratified sample design. The first-stage sampling units consisted of individual schools having 15-year-old students. Schools were sampled systematically from a comprehensive national list of all eligible schools – the school sampling frame – with probabilities that were proportional to a measure of size. This is referred to as systematic probability proportional to size (or PPS) sampling. The measure of size was a function of the estimated number of eligible 15-year-old students enrolled. Prior to sampling, schools in the sampling frame were assigned to mutually exclusive groups called explicit strata, formed to improve the precision of sample-based estimates. The second-stage sampling units in countries using the two-stage design were students within sampled schools. Once schools were selected to be in the sample, a 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 35 although with agreement countries could use alternative values. From each list of students that contained more than the *TCS*, the *TCS* students were selected with equal probability and for lists of fewer than the *TCS*, all students on the list were selected.

In one country, a three-stage design was used. In this case, geographical areas were sampled first (first-stage units) using probability proportional to size sampling, and then schools (second-stage units) were selected within sampled areas. Students were the third-stage sampling units in three-stage designs.

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.

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

Coverage of the PISA international target population

NPMs might find 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. In an international survey in education, the types of exclusion must be defined internationally 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 Technical Standards.[1]

Exclusion can take place at the school level (the whole school is excluded) or at the within-school level. Areas deemed by the PGB to be part of a country (for the purpose of PISA), but which were not included for sampling, were designated as non-covered areas, and documented as such – although this occurred infrequently. Care was taken in this regard because, when such situations did occur, the national desired target population differed from the international desired target population.

International within-school exclusion rules for students were specified as follows:

- Intellectually disabled students are students who have a mental or emotional disability and who, in the professional opinion of qualified staff, are cognitively delayed such that they cannot perform in the PISA testing situation. This category includes students who are 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 are students who are permanently physically disabled in such a way that they cannot perform in the PISA testing situation. Functionally disabled students who could respond were to be included in the testing;
- Students with insufficient assessment language experience are students who need to meet all of the following criteria: a) are not native speakers of the assessment language(s), b) have limited proficiency in the assessment language(s), and c) have received less than one year of instruction in the assessment language(s). Students with insufficient assessment language experience could be excluded;
- Not assessable for some other reason as agreed upon. A nationally-defined within-school exclusion category was permitted if agreed upon by the consortium. A specific subgroup of students (dyslexic, for example) 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.

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

It was required that the overall exclusion rate within a country be kept below 5%. Restrictions on the level of exclusions of various types were as follows:

- School-level exclusions for inaccessibility, feasibility or other reasons were required to cover fewer than 0.5% of the total number of students in the International Target Population. Schools on the school sampling frame which had only one or two 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 schools would not cause a breach of the 0.5% allowable limit, then such schools could be excluded in the field if at that time, they still only had 1 or 2 PISA eligible students;
- School-level exclusions for intellectually or functionally disabled students, or students with insufficient assessment language experience, were required to cover fewer than 2% of students;
- Because definitions of within-school exclusions could vary from country to country, NPMs were asked to adapt the international definitions to make them workable in their country but still to code them according to the PISA international coding scheme. Within-school exclusions for intellectually disabled or functionally disabled students, or students with insufficient assessment language experience, or students nationally-defined and agreed upon were expected to cover fewer than 2.5% of students. Initially, this could only be an estimate. If the actual percentage was ultimately greater than 2.5%, the percentage was re-calculated without considering students excluded because of insufficient assessment language experience since this is a largely unpredictable part of each country's eligible population, not under the control of the education system. If the resulting percentage was below 2.5%, the exclusions were regarded as acceptable.

Accuracy and precision

A minimum of 150 schools (or all schools if there were fewer than 150 schools in a participating country) had to be selected in each country. Within each participating school, a predetermined number of students, denoted as *TCS* (usually 35), were randomly selected with equal probability, or in schools with fewer than *TCS* eligible students, all students were selected. In total, a minimum sample size of 4 500 assessed students was to be achieved, or the full population if it was less than this size. It was possible to negotiate a TCS that differed from 35, but if it was reduced then the sample size of schools was increased beyond 150, so as to ensure that at least 4 500 students would be assessed. The *TCS* selected per school had to be at least 20, so as to ensure adequate accuracy in estimating variance components within and between schools – a major analytical objective of PISA.

NPMs were strongly encouraged to identify stratification variables to reduce the sampling variance.

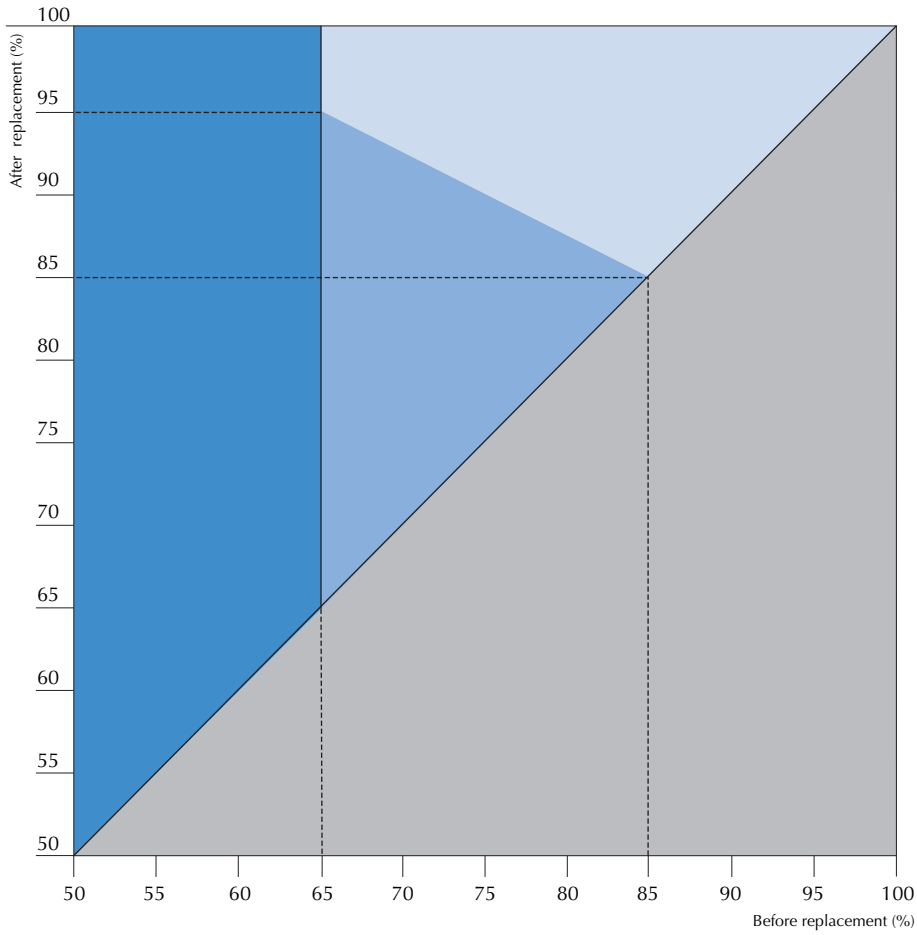
For countries that had participated in PISA 2003 that had larger than anticipated sampling variances associated with their estimates, recommendations were made about sample design changes that would help to reduce the sampling variances for PISA 2006. These included modifications to stratification variables, and increases in the required sample size.

School response rates

A response rate of 85% was required for initially selected schools. If the initial school response rate fell between 65 and 85%, an acceptable school response rate could still be achieved through the use of replacement schools. Figure 4.1 provides a summary of the international requirements for school response rates. To compensate for a sampled school that did not participate, where possible, two potential replacement schools were identified. 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. However, data from such schools were included in the database and contributed to the estimates included in the initial PISA 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.

The rationale for this approach was as follows. 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 50%. 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 be introduced to the extent that the students who were absent were different in achievement from those who attended the testing session, and such a bias is 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 is increased because of sample size attrition.

Figure 4.1
School response rate standard



PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

The judgement was made that, for a school with between 25 and 50% student response, the latter source of bias and variance was likely to introduce more error into the study estimates than the former, but with the converse judgement for those schools with a student response rate below 25%. Clearly the cut-off of 25% is an arbitrary one, 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 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 rapidly decrease.

These PISA 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 in general selected with probability proportional to size, in most countries most schools contributed equal weights, or approximately so, as a consequence the weighted and unweighted school response rates were very similar. Exceptions could occur in countries that had explicit strata that were sampled at very different rates. Details as to how the PISA participants performed relative to these school response rate standards are included in Chapters 10 and 13.

Student response rates

A 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 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% response rate. Again, weighted student response rates were used for assessing this standard. Each student was weighted by the reciprocal of his/her sample selection probability.

MAIN STUDY SCHOOL SAMPLE

Definition of the national target population

NPMs were first required to confirm their dates of testing and age definition with the PISA consortium. 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 define and describe their 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 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. Schools with students that would all be excluded through the within-school exclusion categories could be excluded up to a maximum of 2% as previously noted. Otherwise, countries were instructed to include the schools but to administer the PISA UH booklet, consisting of a subset of the PISA assessment items, deemed more suitable for students with special education needs.

Within participating schools, all 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* 15-year-olds. 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. 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. Country participant 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 nonresponse. Students who could not take the 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.

The 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 that could have 15-year-old students, even those schools which might later be excluded, or deemed ineligible because they had no 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 very careful in constructing their frames.

All but one country used school-level sampling frames as their first stage of sample selection. The *School Sampling Preparation Manual* indicated that the quality of sampling frames for both two and three-stage designs 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 probability-proportional to size selection probabilities were based on it for both two and three-stage designs. The best *ENR* for PISA would have been the number of currently enrolled 15-year-old students. Current enrolment data, however, were rarely available at the time of 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 tend to be enrolled in two or more grades, and the proportions of students who are 15 in each grade are approximately known, the 15-year-old enrolment can be estimated by applying these proportions to the corresponding grade-level enrolments;
- The grade enrolment of the modal grade for 15-year-olds;
- Total student enrolment, divided by the number of grades in the school.

The *School Sampling Preparation Manual* [3] noted that if reasonable estimates of *ENR* did not exist or if the available enrolment data were too out of date, schools might have to be selected with equal probabilities which might require an increased school sample size. No countries needed 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;
- Coded information about the school, such as region of country, school type and extent of urbanisation, which could be used as stratification variables.

As noted, a three-stage design and an area-level sampling frame could be used where a comprehensive national list of schools was not available and could not be constructed without undue burden, or where the procedures for administering the test required that the schools be selected in geographic clusters. As a consequence, the area-level sampling frame introduced an additional stage of frame creation and sampling (called the first stage of sampling) before actually sampling schools (the second stage of sampling). Although generalities about three-stage sampling and using an area-level sampling frame were outlined in the *School Sampling Preparation Manual* (for example that there should be at least 80 first-stage units and about half of them needed to be sampled), NPMs were also instructed in the *School Sampling Preparation Manual* that the more detailed procedures outlined there for the general two-stage design could easily be adapted to the three-stage design. The NPM using a three-stage design was also asked to notify the consortium and received additional support in using an area-level sampling frame. The only country that used a three-stage design was the Russian Federation, where a national list of schools was not available.

Stratification

Prior to sampling, schools were to be ordered, or stratified, in the sampling frame. Stratification consists of classifying schools into like groups according to some variables – referred to as stratification variables. Stratification in PISA 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;
- Make sure that all parts of a population were included in the sample;
- Ensure adequate representation of specific groups of the target population in the sample.

There were two types of stratification possible: explicit and implicit. Explicit stratification consists of building separate school lists, or sampling frames, according to the set of explicit stratification variables under consideration. Implicit stratification consists essentially of sorting the schools within each explicit stratum by a set of implicit stratification variables. This type of stratification is a very simple way of ensuring a strictly proportional sample allocation of schools across all implicit strata. It can also lead to improved reliability of survey estimates, provided that the implicit stratification variables being considered are correlated with PISA achievement (at the school level). Guidelines were provided in the *School Sampling Preparation Manual* on how to go about choosing stratification variables.

Table 4.1 provides the explicit stratification variables used by each country, as well as the number of explicit strata, and the variables and their number of levels used for implicit stratification. As countries were requested to sort the sampling frame by school size, school size was also an implicit stratification variable, though it is not listed in Table 4.1. A variable used for stratification purposes is not necessarily included in the PISA data files.

Table 4.1 [Part 1/2]
Stratification variables

	Explicit stratification variables	Number of explicit strata	Implicit stratification variables
OECD Australia	State/Territory (8); Sector (3); School Size (3)	37	Geographic Zone (8); School Level for TAS and ACT Government Schools (3)
Austria	Programme (19); School Size (3)	20	Province-District (121)
Belgium			
Belgium (Flanders)	Form of Education (5); Public/Private (2); School Size (3);	11	Index of Over-aged Students
Belgium (French)	Special Education/Other (2); Public/Private School Types for Regular Schools (4)	5	Public/Private School Types for Special Education Schools (4); Index of Over-aged Students for Regular Schools
Belgium (German)	One Explicit Stratum (All of German Belgium)	1	None
Canada	Province (10); Language (3); School Size (4); Certainty Selections	44	Public/Private(2); Urban/Rural(2)
Czech Republic	Programmes (6); Region for Programmes 1 and 2 (14); School Size (4)	76	Region for Programmes 3, 4, 5, 6 (14)
Denmark	School Size (3)	3	School Type (5); Geo Area (5)
Finland	Region (6); Urban/Rural (2)	12	None
France	School Type (4); School Size (3)	6	None
Germany	School Category (3); State (16) for Normal Schools; School Size (3)	20	School Type for Normal Schools (5); State for Other Schools (16)
Greece	Region (10); Public/Private (2); Evening/Non-Evening (2); School Size (3)	16	School Type (3); Public/Private (2) when both in an explicit stratum
Hungary	School Type (4); School Size (3)	5	Region (7); National Grade 10 Math Score Categories (5) for Non-Primary Schools
Iceland	Region (9)	9	Urban/Rural (2); School Size (4)
Ireland	School Size (3)	3	School Type (3); School Gender Composition Categories (5)
Italy	Area (17); Programme (5); School Size (3); Certainty Selections	87	Public/Private (2)
Japan	Public/Private (2); School Type (2)	4	Levels of proportion of students taking University/College Entrance Exams(4)
Luxembourg	School Type (6)	6	None
Mexico	State (32); School Level (2); School Size (3); Certainty Selections	67	School Size (3); Public/Private (2); Urban/Rural (2); School Level (2); School Program (4 For Each School Level)
Netherlands	School Track (2)	2	School Type (6 for School Track A and 3 for School Track B)
New Zealand	Certainty/Non-Certainty (2)	2	Public/Private (2); Socio-Economic Status Category (3) and Urban/Rural (2) for Public Schools
Norway	School Type (2); Size (3)	4	None
Poland	Public Upper Secondary Lycea/Other Public (2); School Size (3) for Private Schools	5	Urbanisation (4)
Portugal	Region (7); School Type (4); School Size (3); Certainty Selections	27	Public/Private (2); Socio-Economic Status Category (4)
Korea	Urbanicity (3); School Program (3); School Size (2)	5	School Level (2)
Scotland	School S-Grade Attainment (5)	5	None
	Certainty Selections (1)	2	
Slovak Republic	Region (8); School Type (3); School Size (3)	26	Educational Programme (9); Language (2) in 4 of the Regions
Spain	Region (18); Public/Private (2); Teaching Modality for Basque (3); School Size (4); Certainty Selections	55	Postal Code for all
Sweden	School Size (2); Public/Private (2) for Lower Secondary schools; Urbanicity (5) for large Public Lower Secondary Schools; School Level (2)	10	Urbanisation (5) for Private Lower Secondary schools; Public/private (2) for Upper Secondary schools; Administrative Province (25) for Upper Secondary schools; Income Quartiles (4) for all except Private Lower Secondary schools
Switzerland	School has Grade 9 or not (2); Language (3); School Type (28) for Upper Secondary Schools; Public/Private (2); School Size (4); Certainty Selections	48	School Type (28); Canton (26)
Turkey	Regions (7); School Size (2); Certainty Selections	9	School Level (3); Public/Private (2); Urban/Rural (2)
United Kingdom	PRU/Non-PRU (2), Country (3), Certainty Selections (1)	10	School Type (6), GCSE Performance (6), Region (4), Local Authority, Education and Library Board Region (5)
		2	
		2	
United States	School Size (2)	2	Public/Private (2); Region (4); Urbanisation (3); Minority Status (2); Grade Span (4); Postal Code

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

Table 4.1 [Part 2/2]
Stratification variables

		Explicit stratification variables	Number of explicit strata	Implicit stratification variables
Partners	Argentina	Province (24); School Size (3)	27	Sector (2); School Type (5); School Level (5)
	Azerbaijan	Language (2); Public/Private(2); Education Department (3); School Type (4); School Size (3)	9	Urbanisation (4); Education Department (5)
	Brazil	State (27); School Size (3); Certainty Selections	30	School Type (3); HDI Category (2); School Size (3); Urban/Rural (2); Capital/Non-Capital (2)
	Bulgaria	Region (11); School Size (3)	13	School Type (3); Settlement Size (5); State/Municipal (2); Public/Private (2)
	Chile	School Administration (3); School Level (7); School Size (3);	17	Urban/Rural (2); Region (13)
	Colombia	School Size (3)	3	Urban/Rural (2); Public/Private(2)
	Croatia	Dominant Programme(6); Urbanicity (3); School Size (2); Primary Schools (1); Certainty Selections	16	County (21)
	Estonia	Language (2); School Size (3); Certainty Selections	6	Urbanisation (4); School Type (4); County (15)
	Hong Kong-China	School Type (4)	4	Student Academic Intake (4)
	Indonesia	Provinces (26); School Size (3)	28	School Type (5); Public/Private (2); National Achievement Score Categories (3)
	Israel	Inspection (5); School Size (3)	9	Location (9) for Public Schools, Except For Schools in Druz Migzar Sector; Group Size (5) for Regular Public Schools; Gender Composition (3) for Religious Public Schools; Migzar Sector (3) for Regular Public Arabic Schools; Cultivation Categories (4) for
	Jordan	School Type (4); School Size (3); Certainty Selections	6	Urbanisation (2); School Gender Composition (3); School form (2)
	Kyrgyzstan	Regions (9); Urbanicity (3); Language (3); School Size (3); Certainty Selections	43	School Type (5)
	Latvia	School Size (3); Certainty Selections	4	Urbanisation (4); School Type (4)
	Liechtenstein	One Explicit Stratum (All of Liechtenstein)	1	None
	Lithuania	School Type (4); School Size (3)	6	Urbanisation (3); Public/Private(2)
	Macao-China	School Type (3); Programme (2); Language (5)	10	Secondary Levels (3)
	Montenegro	Primary/Secondary (2); Region (3) for Secondary Schools	4	Region (3) for Primary Schools; Urban/Rural (2); School Type (3)
	Qatar	School Type (7); School Gender Composition Categories (3); School Level (3)	26	Qatari/Non-Qatari (2)
	Romania	School Program (3); School Size (3)	5	Language (3); Urbanisation (2)
	Russian Federation	Region PSU (45)	45	Urbanisation (9); School Type (4); School Sub-Type (16)
	Serbia	Region (8); School Size (2); Certainty Selections	10	Urban/Rural (2); School Type (7)
	Slovenia	School Programme (6); School Size (2); Certainty Selections	8	Urbanisation (4)
	Chinese Taipei	Region (6); School Type (7); School Size (3); Certainty Selections	17	Public/Private (2)
	Thailand	Department (6); School Type (3); School Size (3); Certainty Selections	12	Local Area (9)
	Tunisia	Public/Private (2); School Type (2); For General Public Schools: East/West (2) and School Level (3); School Size (2) for all; Certainty Selections	10	Category of Grade Repeating (3) for General Public Schools; East/West (2) for Private Schools and Vocational Schools; North/South (2) for all
	Uruguay	School Type (4); Programme (3 or 5 Depending on School Type); School Size (3); Certainty Selections	16	Area (4); Shift (4) for Public Secondary Schools; Shift (4) for Public Technical Schools

Treatment of small schools in stratification

In PISA schools were classified as very small, moderately small or large. A school was classified as large if it had an *ENR* above the *TCS* (35 in most countries). A very small school had an *ENR* less than one-half the *TCS* (17 or less in most countries)**.** A moderately small school had an *ENR* in the range of one-half the *TCS* to *TCS* (17 to 35 in most countries). Unless they received special treatment in the sampling, the occurrence of small schools in the sample will reduce the sample size of students for the national sample to below the desired target because the in-school sample size would fall short of expectations. A sample with many small schools could also be an administrative burden. To minimise these problems, procedures for stratifying and allocating school samples were devised for small schools in the sampling frame.

To determine what was needed – a single stratum of small schools (very small and moderately small combined), or a stratum of very small schools only, or two strata, one of very small schools and one of moderately small schools, or no small school strata – the *School Sampling Preparation Manual* stipulated that if:

- The percentage of students in very small schools was 1% or more and the percentage of students in moderately small schools was 4% or more, then an explicit stratum of moderately small schools and an explicit stratum for very small schools were required;
- Otherwise, if the percentage of students in very small schools was 1% or more, a stratum for very small schools was needed, but no stratum for moderately small schools;
- Otherwise, if the percentage of students in very small schools was less than 1%, and the percentage of students in moderately small schools was 4% or more, a combined stratum for small schools which included all very small and moderately small schools, was needed;
- Otherwise no small school strata were needed.

The small school strata were always sorted first by the explicit stratum to which they originally belonged, followed by the other defined implicit stratification variables.

When small schools were explicitly stratified, it was important to ensure that an adequate sample was selected without selecting too many small schools as this would lead to too few students in the assessment. In this case, the entire school sample would have to be increased to meet the target student sample size.

The sample had to be proportional to the number of students and not to the number of schools. Suppose that 10% of students attend moderately small schools, 10% very small schools and the remaining 80% attend large schools. In the sample of 5 250, 4 200 students would be expected to come from large schools (*i.e.* 120 schools with 35 students), 525 students from moderately small schools and 525 students from very small schools. If moderately small schools had an average of 25 students, then it would be necessary to include 21 moderately small schools in the sample. If the average size of very small schools was 10 students, then 52 very small schools would be needed in the sample and the school sample size would be equal to 193 schools rather than 150.

To balance the two objectives of selecting an adequate sample of explicitly stratified small schools, a procedure was recommended that assumes identifying strata of both very small and moderately small schools. The underlying idea is to under-sample by a factor of two the very small school stratum and to increase proportionally the sizes of the large school strata. When there was just a single small school stratum, the procedure was modified by ignoring the parts concerning very small schools. The formulae below assume a target school sample size of 150 and a target student sample size of 5 250.

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

- Step 1: From the complete sampling frame, find the proportions of total *ENR* that come from very small schools (*P*), moderately small schools (*Q*), and larger schools (those with *ENR* of at least *TCS*) (*R*). Thus, $P + Q + R = 1$.
- Step 2: Calculate the figure *L*, where $L = 1.0 + (P/2)$. Thus *L* is a positive number slightly more than 1.0.
- Step 3: The minimum sample size for larger schools is equal to 150 × R × L, rounded 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 (*VENR*). *MENR* is a number in the range of *TCS/2* to *TCS*, and *VENR* is a number no greater than *TCS/2*.
- Step 5: The number of schools that must be sampled from the stratum of moderately small schools is given by: $(5\,250 \times Q \times L)/(MENR)$.
- Step 6: The number of schools that must be sampled from the stratum of very small schools is given by: $(2\,625 \times P \times L)/(VENR)$.

To illustrate the steps, suppose that in participant country *X*, the *TCS* is equal to 35, with 0.1 of the total enrolment of 15-year-olds each in moderately small schools and in very small schools. Suppose that the average enrolment in moderately small schools is 25 students, and in very small schools it is 10 students. Thus $P = 0.1$, $Q = 0.1$, $R = 0.8$, $MENR = 25$ and $VENR = 10$.

From Step 2, $L = 1.05$, then (*Step 3*) the sample size of larger schools must be at least $150 \times (0.80 \times 1.05) = 126$. That is, at least 126 of the larger schools must be sampled. From Step 5, the number of moderately small schools required is $(5\,250 \times 0.1 \times 1.05)/25 = 22.1$ – *i.e.*, 22 schools. From Step 6, the number of very small schools required is $(2\,625 \times 0.1 \times 1.05)/10 = 27.6$ – *i.e.*, 28 schools.

This gives a total sample size of 126 + 22 + 28 = 176 schools, rather than just 150, or 193 as calculated above. Before considering school and student non-response, the larger schools will yield a sample of $126 \times 35 = 4\,410$ students. The moderately small schools will give an initial sample of approximately $22 \times 25 = 550$ students, and very small schools will give an initial sample size of approximately $28 \times 10 = 280$ students. The total initial sample size of students is therefore 4 410 + 550 + 280 = 5 240.

Assigning a measure of size to each school

For the probability proportional to size sampling method used for PISA, a measure of size (*MOS*) derived from *ENR* was established for each school on the sampling frame. *MOS* was constructed as: *MOS* = max (*ENR, TCS*).

The measure of size was therefore equal to the enrolment estimate, unless it was less than the *TCS*, in which case it was set equal to the target cluster size. In most countries, the *MOS* was equal to *ENR* or 35, whichever was larger.

As sample schools were selected according to their size (PPS), setting the measure of size of small schools to 35 is equivalent to drawing a simple random sample of small schools.

School sample selection

Sorting the sampling frame

The *School Sampling Preparation Manual* indicated that, prior to selecting schools from the school sampling frame, schools in each explicit stratum were to be sorted by 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 sorting variable, and so on, until all implicit stratification variables were exhausted. 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, etc., through all implicit strata within an explicit stratum.

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 eligible students in each corresponding explicit stratum. There were two exceptions. If an explicit stratum of very small schools was required, students in them had smaller percentages in the sample than those in the population. To compensate for the resulting loss of sample, the large school strata 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 these cases, two schools were allocated for selection in the stratum to aid with variance estimation.

Determining which schools to sample

The PPS-systematic sampling method used in PISA 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: $I = S/D$;
- Recording the sampling interval, *I*, to four decimal places.

Next, a random number (drawn from a uniform distribution) had to be selected for each explicit stratum**.** The generated random number (*RN*) was to be a number between 0 and 1 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 first selection number was used to identify the first sampled school in the specified explicit stratum;
- Obtaining the second selection number by simply 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.

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

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

PISA and TIMSS or PIRLS overlap control

The main studies for PISA 2006 and the 2007 Trends in International Mathematics and Science Study (TIMSS) were to occur at approximately the same time in southern hemisphere countries and in northern hemisphere countries with late PISA testing. Furthermore, the PISA 2006 main study and the 2006 Progress in International Reading Literacy Study (PIRLS) were to occur at approximately the same time. Because of the potential for increased burden, an overlap control procedure was used for eight countries (Australia, Bulgaria, England, Hong Kong-China, Hungary, Scotland, Tunisia, and the USA) who wished for there to be a minimum incidence of the same schools being sampled for PISA and TIMSS (Australia, Bulgaria, England, Hong Kong-China, Scotland, Tunisia, and the USA) or a minimum of the same schools for PISA and PIRLS (Hungary). This overlap control procedure required that the same school identifiers be used on the TIMSS or PIRLS and PISA school frames for the schools in common.

The TIMSS and PIRLS samples were selected before the PISA samples. Thus, for countries requesting overlap control, the TIMSS and PIRLS International Study Center supplied the PISA consortium with their school frames, with the school IDs, the school probability of selection for each school, and an indicator showing which schools had been sampled for the relevant study.

Sample selections for PISA and the other study could totally avoid overlap of schools if schools which would have been selected with high probability for either study had their selection probabilities capped at 0.5. Such an action would make each study's sample slightly less than optimal, but this might be deemed acceptable when weighed against the possibility of low response rates due to school burden. Each study's project manager had to decide if this was the path they wished to adopt. If they decided against this capping of probabilities, then it might have been possible for some large schools to be in both the PISA and the other study's samples. Among the countries choosing overlap control in the 2006 PISA, selection probabilities were capped at 0.5 only for Hong Kong-China. In the other countries, if any schools had probabilities of selection greater than 0.5 on either study frame, these schools had the possibility to be selected to be in both studies.

To control overlap, the sample selection of schools for PISA adopted a modification of the approach due to Keyfitz (1951), based on Bayes Theorem. To use TIMSS and PISA in an example of the overlap control approach, suppose that *PROBT* is the TIMSS probability of selection, and *PROBP* is the required PISA probability of selection. Then a conditional probability of selection into PISA, *CPROB* is determined as follows:

4.1

$$CPROB = \begin{cases} \max\left[0, \left(\dfrac{PROBT + PROBP - 1}{PROBT}\right)\right] & \text{if the school was TIMSS selected} \\ \min\left[1, \dfrac{PROBP}{(1 - PROBT)}\right] & \text{if the school was not TIMSS selected} \\ PROBP & \text{if the school was not a TIMSS eligible school} \end{cases}$$

Then a conditional *MOS* variable was created to coincide with these conditional probabilities as follows:

CMOS=*CPROB* × stratum sampling interval (recorded to 4 decimal places)

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

The PISA school sample was then selected using the line numbers created as usual (see below), but applied to the cumulated *CMOS* values (as opposed to the cumulated *MOS* values). Note that it was possible that the resulting PISA sample size could be slightly lower or higher than the originally assigned sample size, but this was deemed acceptable.

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 determined 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.

Identifying replacement schools

Each sampled school in the main survey was assigned two replacement schools from the sampling frame, identified as follows. For each sampled school, the schools immediately preceding and following it in the explicit stratum 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. The *School Sampling Preparation Manual* noted that in small countries, there could be problems when trying to identify two replacement schools for each sampled school. In such cases, a replacement school was allowed to be the potential replacement for two sampled schools (a first replacement for the preceding school, and a second replacement for the following school), but an actual replacement for only one school. Additionally, it may have been difficult to assign replacement schools for some very large sampled schools because the sampled schools appeared very close to each other in the sampling frame. There were times when it was only possible to assign a single replacement school, or even none, when two consecutive schools in the sampling frame were sampled.

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, in which case the two schools immediately following it were designated as replacement schools.

Assigning school identifiers

To keep track of sampled and replacement schools in the PISA database, each was assigned a unique, three-digit school code and two-digit stratum code (corresponding to the explicit strata) sequentially numbered starting with one within each explicit stratum. For example, if 150 schools are sampled from a single explicit stratum, they are assigned identifiers from 001 to 150. First replacement schools in the main survey are assigned the school identifier of their corresponding sampled schools, incremented by 300. For example, the first replacement school for sampled school 023 is assigned school identifier 323. Second replacement schools in the main survey are assigned the school identifier of their corresponding sampled schools, but incremented by 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. They contacted replacement schools after all contacts with sampled schools were made. Each sampled school that did not participate was replaced if possible. If both an original school and a replacement participated, only the data from the original school were included in the weighted data provided that at least 50% of the eligible, non-excluded students had participated. If this was not the case, it was permissible for the original school to be labelled as a nonrespondent and the replacement school as the respondent, provided that the replacement school had at least 50% of the eligible, non-excluded students as participants.

Monitoring school sampling

For PISA 2006, it was a strong recommendation that the consortium select the school samples. This was incorporated into the 2006 procedures to alleviate the weighting difficulties caused by receiving school frame files in many different formats. France and Japan selected their own school samples for reasons of confidentiality. Sample selection was replicated by the consortium to ensure quality. All other samples were selected by and checked in detail by the consortium. All countries were required to submit sampling forms 1 (time of testing and age definition), 2 (national desired target population), 3 (national defined target population), 4 (sampling frame description), 5 (excluded schools), 7 (stratification), and 11 (school sampling frame). The consortium completed and returned the others (forms 6, 8, 9, 10, and the base form 12) for countries for which they did the sampling. Otherwise, the country also submitted these other forms for approval. Table 4.2 provides a summary of the information required on each form and the timetables (which depended on national assessment periods).

Table 4.2
Schedule of school sampling activities

Activity	Submit to Consortium	Due Date
Specify time of testing and age definition of population to be tested	Sampling form 1 – time of testing and age definition	Submit three months before the school sample is to be selected
Define national desired target population	Sampling form 2 – national desired target population	Submit three months before the school sample is to be selected
Define national defined target population	Sampling form 3 – national defined target population	Submit three months before the school sample is to be selected
Create and describe sampling frame	Sampling form 4 – sampling frame Description	Submit two months before the school sample is to be selected
Decide on schools to be excluded from sampling frame	Sampling form 5 – excluded schools	Submit two months before the school sample is to be selected
Decide how to treat small schools	Sampling form 6 – Treatment of Small schools	The Consortium will complete and return this form to the NPM about one month before the school sample is to be selected.
Decide on explicit and implicit stratification variables	Sampling form 7 – stratification	Submit two months before the school sample is to be selected
Describe population within strata	Sampling form 8 – population counts by strata	The Consortium will complete and return this form to the NPM when the school sample is sent to the NPM.
Allocate sample over explicit strata	Sampling form 9 – sample allocation by explicit strata	The Consortium will complete and return this form to the NPM about one month before the school sample is to be selected.
Select the school sample	Sampling form 10 – school sample Selection	The Consortium will complete and return this form to the NPM when the school sample is sent to the NPM.
Identify sampled schools, replacement schools and assign PISA school IDs	Sampling form 11 – school sampling frame	Submit two months before the school sample is to be selected. The Consortium will return this form to the NPM with sampled schools and their replacement schools identified and with PISA IDs assigned when the school sample is selected.
Create a school tracking form	Sampling form 12 – school tracking form	Submit within one month of the end of the data collection period

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

Once received from each country, each form 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 consortium. In cases where approval could not be granted, countries were asked to make revisions to their sample design and sampling forms.

Checks that were performed in the monitoring of each form follow. All entries were observed in their own right but those below are additional matters explicitly examined.

Sampling form 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;
- Assessment dates were checked against recorded main study (MS) assessment dates on field trial (FT) sampling forms. Differences were queried;
- If assessment end dates were close to the end of the population birth date window, NPMs were alerted not to conduct any make-up sessions beyond the date when the population births dates were valid;
- Population birth dates were checked against those recorded for the MS on the FT sampling forms. Differences were queried.

Sampling form 2: National desired target population

- Large deviations between the total national number of 15-year-olds and the enrolled number of 15-year-olds were questioned;
- Large increases or decreases in population numbers compared to those from PISA 2003 were queried, as were seeming trends in population numbers (increasing or decreasing) since PISA 2000;
- Any population to be omitted from the international desired population was noted and discussed, especially if the percentage of 15-year-olds to be excluded was more than 2% or if it was not noted for PISA 2003;
- Calculations were verified;
- For any countries using a three-stage design, a sampling form 2 also needed to be completed for the full national desired population as well as for the population in the sampled regions;
- For countries having adjudicated regions, a sampling form 2 was needed for each region;

Sampling form 3: National defined target population

- The population figure in the first question needed to correspond with the final population figure on sampling form 2;
- Reasons for excluding schools were checked for appropriateness;
- Exclusion types and extents were compared to those recorded for PISA 2003. Differences were queried;
- Use of the UH booklet was queried;
- Exclusions for language were checked against what was recorded for the MS on the FT sampling forms. Differences were queried;
- The number and percentage of students to be excluded at the school level and whether the percentage was less than the maximum percentage allowed for such exclusions were checked;

- Calculations were verified and the overall coverage figures were assessed;
- Reasonableness of assumptions about within-school exclusions was assessed by checking previous PISA coverage tables;
- The population figures on this form were compared against the summed sampling frame enrolment. Differences were queried;
- For any countries using a three-stage design, a sampling form 3 also needed to be completed for the full national defined population as well as for the population in the sampled regions;
- For countries having adjudicated regions, a sampling form 3 was needed for each region.

Sampling form 4: Sampling frame description

- Special attention was paid to countries who reported on this form that a three-stage sampling design was to be implemented and additional information was sought from countries in such cases to ensure that the first-stage sampling was done adequately;
- The type of school-level enrolment estimate and the year of data availability were assessed for reasonableness;
- Frame sampling units were compared against those used for PISA 2003. Differences were queried.

Sampling form 5: Excluded schools

- The number of schools and the total enrolment figures, as well as the reasons for exclusion, were checked to ensure correspondence with figures reported on sampling form 3 about school-level exclusions.

Sampling form 6: Treatment of small schools

- Calculations were verified, as was the decision about whether or not a moderately small schools stratum and/or a very small schools stratum were needed.

Sampling form 7: Stratification

- Since explicit strata are formed 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, each country's choice of explicit stratification variables was assessed. If a country was known to have school tracking, and tracks or school programmes were not among the explicit stratifiers, a suggestion was made to include this type of variable;
- Identified stratification variables were compared against those noted for the MS on the FT sampling forms. Differences were queried;
- 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;
- The sampling frame was checked to ensure that the stratification variables were available for all schools. Different explicit strata were allowed to have different implicit stratifiers;
- Any indicated student sorting variables were compared to those used in PISA 2003. Differences were queried.

Sampling form 8: Population counts by strata

- Counts on sampling form 8 were compared to counts arising from the frame. Any differences were queried and corrected as appropriate.

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

Sampling form 9: Sample allocation by explicit strata

- All explicit strata had to be accounted for on sampling form 9;
- All explicit strata population entries were compared to those determined from the sampling frame;
- 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;
- The percentage of students in the sample for each explicit stratum had to be close to the percentage in the population for each stratum (very small schools strata were an exception since under-sampling was allowed);
- 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;
- Previous PISA response rates were reviewed and if deemed necessary, sample size increases were suggested.

Sampling form 10: School sample selection

- All calculations were verified;
- Particular attention was paid to the four decimal places that were required for both the sampling interval and the random number.

Sampling form 11: School sampling frame

- The frame number of sampling units was compared to the same for PISA 2003. Differences were queried;
- NPMs were queried about whether or not they had included schools with grades 7 or 8 that could potentially have PISA students at the time of assessment;
- NPMs were queried about whether or not they had included vocational or apprenticeship, schools with only part-time students, International or foreign schools or any other irregular schools that could contain PISA students at the time of the assessment;
- The frame was checked for proper sorting according to the implicit stratification scheme and enrolment values, and the proper assignment of the measure of size value, especially for moderately small and very small schools. The accumulation of the measure of size values was also checked for each explicit stratum. This final cumulated measure of size value for each stratum had to correspond to the 'Total Measure of Size' value on sampling form 10 for each explicit stratum. Additionally, each line selection number was checked against the frame cumulative measure of size figures to ensure that the correct schools were sampled. Finally, the assignment of replacement schools and PISA identification numbers were checked to ensure that all rules laid out in the *Sampling Manual* were adhered to. Any deviations were discussed with each country and either corrected or the deviations accepted.

Sampling form 12: School tracking form

- Sampling form 12 was checked to see that the PISA identification numbers on this form matched those on the sampling frame;
- Checks were made to ensure that all sampled and replacement schools were accounted for;
- Checks were also made to ensure that status entries were in the requested format.

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

Student samples

Student selection procedures in the main study were the same as those used in the field trial. Student sampling was generally undertaken using the consortium software, *KeyQuest*, at the national centres from lists of all eligible students in each school that had agreed to participate. 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 very important that the student sample be selected from accurate, complete lists, the lists needed to be prepared not too far in advance of the testing and had to list all eligible students. It was suggested that the lists be received one to two months before testing so that the NPM would have the time to select the student samples.

Twelve countries (Chile, the Czech Republic, Germany, Iceland, Japan, Korea, Liechtenstein, Mexico, Norway, Sweden, Switzerland and Uruguay) chose student samples that included students aged 15 and/or enrolled in a specific grade (*e.g.*, grade 10). Thus, a larger overall sample, including 15-year-old students and students in the designated grade (who may or may not have been aged 15) was selected. The necessary steps in selecting larger samples are noted where appropriate in the following steps. The Czech Republic, Korea, Mexico, Norway, Sweden, Switzerland (only in some explicit strata), and Uruguay used the standard method of direct student sampling described here. However, Mexico also sub-sampled schools in which to do the grade sampling from its large school sample. For Iceland and Japan, the sample constituted a de facto grade sample because nearly all of the PISA eligible 15-year-olds were in the grade sampled. Germany, Liechtenstein, and Switzerland (in a second set of explicit strata) supplemented the standard method with an additional sample of grade-eligible students which was selected by first selecting grade 9 classes within PISA sampled schools that had this grade. In Chile, the standard method was supplemented with additional grade-eligible students from a sample of grade 10 classes within PISA sampled schools that had this grade.

Preparing a list of age-eligible students

Each school drawing an additional grade sample was to prepare a list of age and grade-eligible students that included all students in the designated grade (*e.g.*, grade 10); and all other 15-year-old students (using the appropriate 12-month age span agreed upon for each country) currently enrolled in other grades. This form was referred to as a student listing form. The following were considered important:

- Age-eligible students were all students born in 1990 (or the appropriate 12-month age span agreed upon for the 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 sample was selected;
- It was suggested that schools retain a copy of the list in case the NPM had to call the school with questions;
- A computer list was to be up-to-date at the time of sampling rather than prepared at the beginning of the school year. Students were identified by their unique student identification numbers.

Selecting the student sample

Once NPMs received the list of eligible students from a school, the student sample was to be selected and the list of selected students (*i.e.* the student tracking form) returned to the school. NPMs were required to use *KeyQuest*, the PISA sampling software, to select the student samples unless agreed upon. Three countries (Germany, Luxembourg, and Switzerland) did not use *KeyQuest* for all or for a part of the student sample

for reasons including extra student demographic data or due to an unusual, but approved, class sampling approach for a grade option.

Preparing instructions for excluding students

PISA was a timed assessment administered in the instructional language(s) of each country and designed to be as inclusive as possible. For students with limited assessment language(s) experience or with physical, mental, or emotional disabilities who could not participate, PISA developed instructions in cases of doubt about whether a selected student should be assessed. NPMs used the guidelines given to develop any additional instructions; school co-ordinators and test administrators needed precise instructions for exclusions. The national operational definitions for within-school exclusions were to be well documented and submitted to the consortium 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 them and their teachers (and parents), to update information and to identify the students to be excluded. The student tracking form was therefore sent about two weeks before the assessment session. It was recommended that a copy of the tracking form be made and kept at the national centre. Another recommendation was to have the NPM send a copy of the form to the test administrator in case the school copy was misplaced before the assessment day. The test administrator and school co-ordinator manuals (see Chapter 6) both assumed that each would have a copy.

In the interest of ensuring PISA was as inclusive as possible, student participation and reasons for exclusion were separately coded in the student tracking form. This allowed for students with special education needs (SEN) to be included when their SEN was not severe enough to be a barrier to their participation. The participation status could therefore show, for example, that a student participated and was not excluded for SEN reasons even though the student was noted with a special education need. Any student whose participation status indicated they were excluded for SEN reasons had to have an SEN code explaining the reason for exclusion. It was important that these criteria be followed strictly for the study to be comparable within and across countries. When in doubt, the student was included. The instructions for excluding students are provided in the PISA Technical Standards.

Notes

1. A student was deemed a participant if they gave at least one response to the cognitive assessment, or they responded to at least one student questionnaire item and either they or their parents provided the occupation of a parent or guardian (*see* Chapter 17).

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

Reader's Guide

Country codes – the following country codes are used in this report:

OECD countries

AUS	Australia
AUT	Austria
BEL	Belgium
BEF	Belgium (French Community)
BEN	Belgium (Flemish Community)
CAN	Canada
CAE	Canada (English Community)
CAF	Canada (French Community)
CZE	Czech Republic
DNK	Denmark
FIN	Finland
FRA	France
DEU	Germany
GRC	Greece
HUN	Hungary
ISL	Iceland
IRL	Ireland
ITA	Italy
JPN	Japan
KOR	Korea
LUX	Luxembourg
LXF	Luxembourg (French Community)
LXG	Luxembourg (German Community)
MEX	Mexico
NLD	Netherlands
NZL	New Zealand
NOR	Norway
POL	Poland
PRT	Portugal
SVK	Slovak Republic
ESP	Spain
ESB	Spain (Basque Community)
ESC	Spain (Catalonian Community)
ESS	Spain (Castillian Community)
SWE	Sweden
CHE	Switzerland
CHF	Switzerland (French Community)
CHG	Switzerland (German Community)
CHI	Switzerland (Italian Community)
TUR	Turkey
GBR	United Kingdom
IRL	Ireland
SCO	Scotland
USA	United States

Partner countries and economies

ARG	Argentina
AZE	Azerbaijan
BGR	Bulgaria
BRA	Brazil
CHL	Chile
COL	Colombia
EST	Estonia
HKG	Hong Kong-China
HRV	Croatia
IDN	Indonesia
JOR	Jordan
KGZ	Kyrgyztan
LIE	Liechtenstein
LTU	Lithuania
LVA	Latvia
LVL	Latvia (Latvian Community)
LVR	Latvia (Russian Community)
MAC	Macao-China
MNE	Montenegro
QAT	Qatar
ROU	Romania
RUS	Russian Federation
SRB	Serbia
SVN	Slovenia
TAP	Chinese Taipei
THA	Thailand
TUN	Tunisia
URY	Uruguay

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

References

Adams, R.J., Wilson, M. & **Wang, W.C.** (1997), The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement,* No. 21, pp. 1-23.

Adams, R.J., Wilson, M. R. & **Wu, M.L.** (1997), Multilevel item response models: An approach to errors in variables regression, *Journal of Educational and Behavioural Statistics,* No. 22 (1), pp. 46-75.

Adams, R.J. & **Wu, M.L.** (2002), *PISA 2000 Technical Report,* OECD, Paris.

Bollen, K.A. & **Long, S.J.** (1993) (eds.), *Testing Structural Equation Models*, Newbury Park: London.

Beaton, A.E. (1987), Implementing the new design: The NAEP 1983-84 technical report (Rep. No. 15-TR-20), Princeton, NJ: Educational Testing Service.

Buchmann, C. (2000), Family structure, parental perceptions and child labor in Kenya: What factors determine who is enrolled in school? *Soc. Forces,* No. 78, pp. 1349-79.

Buchmann, C. (2002), Measuring Family Background in International Studies of Education: Conceptual Issues and Methodological Challenges, in Porter, A.C. and Gamoran, A. (eds.). *Methodological Advances in Cross-National Surveys of Educational Achievement* (pp. 150-97), Washington, DC: National Academy Press.

Creemers, B.P.M. (1994), *The Effective Classroom*, London: Cassell.

Cochran, W.G. (1977), *Sampling techniques,* third edition, New York, NY: John Wiley and Sons.

Ganzeboom, H.B.G., de Graaf, P.M. & **Treiman, D.J.** (1992), A standard international socio-economic index of occupational status, *Social Science Research*, No. 21, pp. 1-56.

Ganzeboom H.B. & **Treiman, D.J.** (1996), Internationally comparable measures of occupational status for the 1988 international standard classification of occupations, *Social Science Research*, No. 25, pp. 201-239.

Grisay, A. (2003), Translation procedures in OECD/PISA 2000 international assessment, *Language Testing,* No. 20 (2), pp. 225-240.

Hambleton, R.K., Swaminathan, H. & **Rogers, H.J.** (1991), *Fundamentals of item response theory*, Newbury Park, London, New Delhi: SAGE Publications.

Hambleton, R.K., Merenda, P.F. & **Spielberger, C.D.** (2005), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment,* IEA Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.

Harkness, J.A., Van de Vijver, F.J.R. & **Mohler, P.Ph** (2003), *Cross-Cultural Survey Methods,* Wiley-Interscience, John Wiley & Sons, Inc., Hoboken, New Jersey.

Harvey-Beavis, A. (2002), Student and School Questionnaire Development, in R.J. Adams and M.L. Wu (eds.), *PISA 2000 Technical Report,* (pp. 33-38), OECD, Paris.

International Labour Organisation (ILO) (1990), *International Standard Classification of Occupations: ISCO-88.* Geneva: International Labour Office.

Jöreskog, K.G. & **Sörbom, Dag** (1993), *LISREL 8 User's Reference Guide*, Chicago: SSI.

Judkins, D.R. (1990), Fay's Method of Variance Estimation, *Journal of Official Statistics*, No. 6 (3), pp. 223-239.

Kaplan, D. (2000), *Structural equation modeling: Foundation and extensions,* Thousand Oaks: SAGE Publications.

Keyfitz, N. (1951), Sampling with probabilities proportionate to science: Adjustment for changes in probabilities, *Journal of the American Statistical Association,* No. 46, American Statistical Association, Alexandria, pp. 105-109.

Kish, L. (1992), Weighting for Unequal, *Pi. Journal of Official Statistics*, No. 8 (2), pp. 183-200.

LISREL (1993), K.G. Jöreskog & D. Sörbom, [computer software], Lincolnwood, IL: Scientific Software International, Inc.

Lohr, S.L. (1999), *Sampling: Design and Analysis*, Duxberry: Pacific Grove.

Macaskill, G., Adams, R.J. & **Wu, M.L.** (1998), Scaling methodology and procedures for the mathematics and science literacy, advanced mathematics and physics scale, in M. Martin and D.L. Kelly, Editors, *Third International Mathematics and Science Study, technical report Volume 3: Implementation and analysis*, Boston College, Chestnut Hill, MA.

Masters, G.N. & **Wright, B.D.** (1997), The Partial Credit Model, in W.J. van der Linden, & R.K. Hambleton (eds.), *Handbook of Modern Item Response Theory* (pp. 101-122), New York/Berlin/Heidelberg: Springer.

Mislevy, R.J. (1991), Randomization-based inference about latent variables from complex samples, *Psychometrika,* No. 56, pp. 177-196.

Mislevy, R.J., Beaton, A., Kaplan, B.A. & **Sheehan, K.** (1992), Estimating population characteristics from sparse matrix samples of item responses, *Journal of Educational Measurement,* No. 29 (2), pp. 133-161.

Mislevy, R.J. & **Sheehan, K.M.** (1987), Marginal estimation procedures, in Beaton, A.E., Editor, 1987. *The NAEP 1983-84 technical report,* National Assessment of Educational Progress, Educational Testing Service, Princeton, pp. 293-360.

Mislevy, R.J. & **Sheehan, K.M.** (1989), Information matrices in latent-variable models, *Journal of Educational Statistics,* No. 14, pp. 335-350.

Mislevy, R.J. & **Sheehan, K.M.** (1989), The role of collateral information about examinees in item parameter estimation, *Psychometrika,* No. 54, pp. 661-679.

Monseur, C. & **Berezner, A.** (2007), The Computation of Equating Errors in International Surveys in Education, *Journal of Applied Measurement,* No. 8 (3), 2007, pp. 323-335.

Monseur, C. (2005), An exploratory alternative approach for student non response weight adjustment, *Studies in Educational Evaluation,* No. 31 (2-3), pp. 129-144.

Muthen, B. & **L. Muthen** (1998), [computer software], *Mplus* Los Angeles, CA: Muthen & Muthen.

Muthen, B., du Toit, S.H.C. & **Spisic, D.** (1997), *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes,* unpublished manuscript.

OECD (1999), *Classifying Educational Programmes. Manual for ISCED-97 Implementation in OECD Countries,* OECD, Paris.

OECD (2003), *Literacy Skills for the World of Tomorrow: Further results from PISA 2000,* OECD, Paris.

OECD (2004), *Learning for Tomorrow's World – First Results from PISA 2003,* OECD, Paris.

OECD (2005), *Technical Report for the OECD Programme for International Student Assessment 2003,* OECD, Paris.

OECD (2006), *Assessing Scientific, Reading and Mathematical Literacy: A framework for PISA 2006,* OECD, Paris.

OECD (2007), *PISA 2006: Science Competencies for Tomorrow's World,* OECD, Paris.

PISA Consortium (2006), *PISA 2006 Main Study Data Management Manual, https://mypisa.acer.edu.au/images/mypisadoc/opmanual/pisa2006_data_management_manual.pdf*

Rasch, G. (1960), Probabilistic models for some intelligence and attainment tests, Copenhagen: Nielsen & Lydiche.

Routitski A. & **Berezner, A.** (2006), Issues influencing the validity of cross-national comparisons of student performance. Data Entry Quality and Parameter Estimation. Paper presented at the Annual Meeting of the American Educational Research Association (AERA) in San Francisco, 7-11 April, *https://mypisa.acer.edu.au/images/mypisadoc/aera06routitsky_berezner.pdf*

Rust, K. (1985), Variance Estimation for Complex Estimators in Sample Surveys, *Journal of Official Statistics,* No. 1, pp. 381-397.

Rust, K.F. & **Rao, J.N.K.** (1996), Variance Estimation for Complex Surveys Using Replication Techniques, *Survey Methods in Medical Research,* No. 5, pp. 283-310.

Shao, J. (1996), Resampling Methods in Sample Surveys (with Discussion), *Statistics,* No. 27, pp. 203-254.

Särndal, C.-E., Swensson, B. & **Wretman, J.** (1992), *Model Assisted Survey Sampling,* New York: Springer-Verlag.

SAS® CALIS (1992), W. Hartmann [computer software], Cary, NC: SAS Institute Inc.

Scheerens, J. (1990), School effectiveness and the development of process indicators of school functioning, *School effectiveness and school improvement,* No. 1, pp. 61-80.

Scheerens, J. & **Bosker, R.J.** (1997), *The Foundations of School Effectiveness,* Oxford: Pergamon.

Schulz, W. (2002), Constructing and Validating the Questionnaire composites, in R.J. Adams and M.L. Wu (eds.), *PISA 2000 Technical Report,* OECD, Paris.

Schulz, W. (2004), Mapping Student Scores to Item Responses, in W. Schulz and H. Sibberns (eds.), *IEA Civic Education Study, Technical Report* (pp. 127-132), Amsterdam: IEA.

Schulz, W. (2006a), *Testing Parameter Invariance for Questionnaire Indices using Confirmatory Factor Analysis and Item Response Theory,* Paper presented at the Annual Meetings of the American Educational Research Association (AERA) in San Francisco, 7-11 April.

Schulz, W. (2006b), *Measuring the socio-economic background of students and its effect on achievement in PISA 2000 and PISA 2003,* Paper presented at the Annual Meetings of the American Educational Research Association (AERA) in San Francisco, 7-11 April.

Thorndike, R.L. (1973), *Reading comprehension in fifteen countries,* New York, Wiley: and Stockholm: Almqvist & Wiksell.

Travers, K.J. & **Westbury, I.** (1989), *The IEA Study of Mathematics I: Analysis of Mathematics Curricula,* Oxford: Pergamon Press.

Travers, K.J., Garden R.A. & **Rosier, M.** (1989), Introduction to the Study, in Robitaille, D. A. and Garden, R. A. (eds), *The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics Curricula,* Oxford: Pergamon Press.

Verhelst, N. (2002), Coder and Marker Reliabilaity Studies, in R.J. Adams & M.L. Wu (eds.), *PISA 2000 Technical Report*. OECD, Paris.

Walberg, H.J. (1984), Improving the productivity of American schools, *Educational Leadership,* No. 41, pp. 19-27.

Walberg, H. (1986), Synthesis of research on teaching, in M. Wittrock (ed.), *Handbook of research on teaching* (pp. 214-229), New York: Macmillan.

Walker, M. (2006), *The choice of Likert or dichotomous items to measure attitudes across culturally distinct countries in international comparative educational research*. Paper presented at the Annual Meetings of the American Educational Research Association (AERA) in San Francisco, 7-11 April.

Walker, M. (2007), Ameliorating Culturally-Based Extreme Response Tendencies To Attitude items, *Journal of Applied Measurement,* No. 8, pp. 267-278.

Warm, T.A. (1989), Weighted Likelihood Estimation of Ability in Item Response Theory, *Psychometrika*, No. 54 (3), pp. 427-450.

Westat (2007), *WesVar® 5.1* Computer software and manual, Rockville, MD: Author (also see *http://www.westat.com/wesvar/*).

Wilson, M. (1994), Comparing Attitude Across Different Cultures: Two Quantitative Approaches to Construct Validity, in M. Wilson (ed.), *Objective measurement II: Theory into practice* (pp. 271-292), Norwood, NJ: Ablex.

Wolter, K.M. (2007), *Introduction to Variance Estimation*. Second edition, Springer: New York.

Wu, M.L., Adams, R.J. & **Wilson, M.R.** (1997), *ConQuest®: Multi-Aspect Test Software* [computer program manual], Camberwell, Vic.: Australian Council for Educational Research.

List of abbreviations – the following abbreviations are used in this report:

ACER	Australian Council for Educational Research
AGFI	Adjusted Goodness-of-Fit Index
BRR	Balanced Repeated Replication
CBAS	Computer Based Assessment of Science
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CITO	National Institute for Educational Measurement, The Netherlands
CIVED	Civic Education Study
DIF	Differential Item Functioning
ENR	Enrolment of 15-year-olds
ESCS	PISA Index of Economic, Social and Cultural Status
ETS	Educational Testing Service
IAEP	International Assessment of Educational Progress
I	Sampling Interval
ICR	Inter-Country Coder Reliability Study
ICT	Information Communication Technology
IEA	International Association for the Evaluation of Educational Achievement
INES	OECD Indicators of Education Systems
IRT	Item Response Theory
ISCED	International Standard Classification of Education
ISCO	International Standard Classification of Occupations
ISEI	International Socio-Economic Index
MENR	Enrolment for moderately small school
MOS	Measure of size
NCQM	National Centre Quality Monitor
NDP	National Desired Population
NEP	National Enrolled Population
NFI	Normed Fit Index
NIER	National Institute for Educational Research, Japan
NNFI	Non-Normed Fit Index
NPM	National Project Manager
OECD	Organisation for Economic Cooperation and Development
PISA	Programme for International Student Assessment
PPS	Probability Proportional to Size
PGB	PISA Governing Board
PQM	PISA Quality Monitor
PSU	Primary Sampling Units
QAS	Questionnaire Adaptations Spreadsheet
RMSEA	Root Mean Square Error of Approximation
RN	Random Number
SC	School Co-ordinator
SE	Standard Error
SD	Standard Deviation
SEM	Structural Equation Modelling
SMEG	Subject Matter Expert Group
SPT	Study Programme Table
TA	Test Administrator
TAG	Technical Advisory Group
TCS	Target Cluster Size
TIMSS	Third International Mathematics and Science Study
TIMSS-R	Third International Mathematics and Science Study – Repeat
VENR	Enrolment for very small schools
WLE	Weighted Likelihood Estimates

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

Table of contents

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

LIST OF BOXES

LIST OF FIGURES

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

LIST OF TABLES

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009

PISA 2006 TECHNICAL REPORT – ISBN 978-92-64-04808-9 – © OECD 2009