



Sample design

Target population and overview of the sampling design	64
Population coverage, and school and student participation rate standards	65
▪ Coverage of the PISA international target population	65
▪ Accuracy and precision	66
▪ School response rates	66
▪ Student response rates	68
Main study school sample	68
▪ Definition of the national target population	68
▪ The sampling frame	69
▪ Stratification	70
▪ Assigning a measure of size to each school	74
▪ School sample selection	74
▪ PISA and TIMSS or PIRLS overlap control	76
▪ Student samples	82



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.

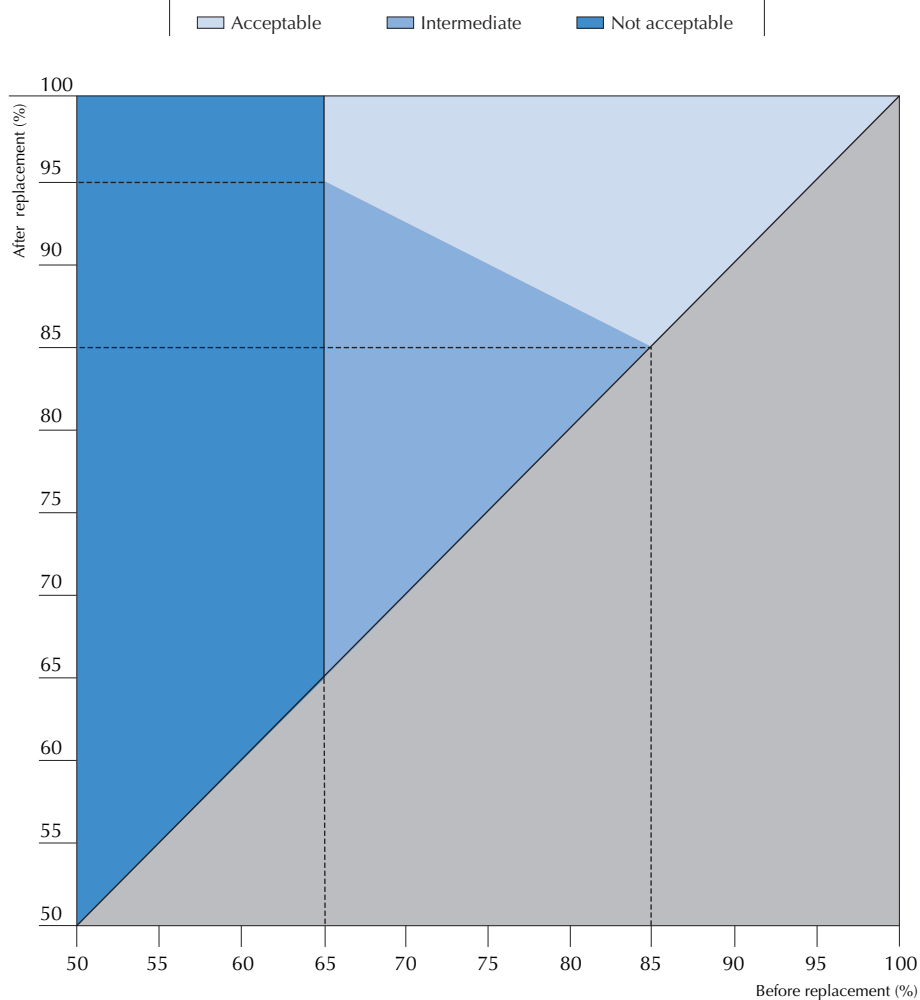
School response rates

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





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	



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)	Sector (2); School Type (5); School Level (5)
	Azerbaijan	Language (2); Public/Private(2); Education Department (3); School Type (4); School Size (3)	Urbanisation (4); Education Department (5)
	Brazil	State (27); School Size (3); Certainty Selections	School Type (3); HDI Category (2); School Size (3); Urban/Rural (2); Capital/Non-Capital (2)
	Bulgaria	Region (11); School Size (3)	School Type (3); Settlement Size (5); State/Municipal (2); Public/Private (2)
	Chile	School Administration (3); School Level (7); School Size (3);	Urban/Rural (2); Region (13)
	Colombia	School Size (3)	Urban/Rural (2); Public/Private(2)
	Croatia	Dominant Programme(6); Urbanicity (3); School Size (2); Primary Schools (1); Certainty Selections	County (21)
	Estonia	Language (2); School Size (3); Certainty Selections	Urbanisation (4); School Type (4); County (15)
	Hong Kong-China	School Type (4)	Student Academic Intake (4)
	Indonesia	Provinces (26); School Size (3)	School Type (5); Public/Private (2); National Achievement Score Categories (3)
	Israel	Inspection (5); School Size (3)	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	Urbanisation (2); School Gender Composition (3); School form (2)
	Kyrgyzstan	Regions (9); Urbanicity (3); Language (3); School Size (3); Certainty Selections	School Type (5)
	Latvia	School Size (3); Certainty Selections	Urbanisation (4); School Type (4)
	Liechtenstein	One Explicit Stratum (All of Liechtenstein)	None
	Lithuania	School Type (4); School Size (3)	Urbanisation (3); Public/Private(2)
	Macao-China	School Type (3); Programme (2); Language (5)	Secondary Levels (3)
	Montenegro	Primary/Secondary (2); Region (3) for Secondary Schools	Region (3) for Primary Schools; Urban/Rural (2); School Type (3)
	Qatar	School Type (7); School Gender Composition Categories (3); School Level (3)	Qatari/Non-Qatari (2)
	Romania	School Program (3); School Size (3)	Language (3); Urbanisation (2)
Russian Federation	Region PSU (45)	Urbanisation (9); School Type (4); School Sub-Type (16)	
Serbia	Region (8); School Size (2); Certainty Selections	Urban/Rural (2); School Type (7)	
Slovenia	School Programme (6); School Size (2); Certainty Selections	Urbanisation (4)	
Chinese Taipei	Region (6); School Type (7); School Size (3); Certainty Selections	Public/Private (2)	
Thailand	Department (6); School Type (3); School Size (3); Certainty Selections	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	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	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.



- 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 $150 \times R \times L$, 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.



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:

$$4.1 \quad CPROB = \begin{cases} \max \left[0, \left(\frac{PROBT + PROBP - 1}{PROBT} \right) \right] & \text{if the school was TIMSS selected} \\ \min \left[1, \frac{PROBP}{(1 - PROBT)} \right] & \text{if the school was not TIMSS selected} \\ PROBP & \text{if the school was not a TIMSS eligible school} \end{cases}$$

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

$CMOS = CPROB \times$ stratum sampling interval (recorded to 4 decimal places)



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



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.



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.



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).



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	Kyrgyzstan
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



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/era06routitski_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 Reliability 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	NPM	National Project Manager
AGFI	Adjusted Goodness-of-Fit Index	OECD	Organisation for Economic Cooperation and Development
BRR	Balanced Repeated Replication	PISA	Programme for International Student Assessment
CBAS	Computer Based Assessment of Science	PPS	Probability Proportional to Size
CFA	Confirmatory Factor Analysis	PGB	PISA Governing Board
CFI	Comparative Fit Index	PQM	PISA Quality Monitor
CITO	National Institute for Educational Measurement, The Netherlands	PSU	Primary Sampling Units
CIVED	Civic Education Study	QAS	Questionnaire Adaptations Spreadsheet
DIF	Differential Item Functioning	RMSEA	Root Mean Square Error of Approximation
ENR	Enrolment of 15-year-olds	RN	Random Number
ESCS	PISA Index of Economic, Social and Cultural Status	SC	School Co-ordinator
ETS	Educational Testing Service	SE	Standard Error
IAEP	International Assessment of Educational Progress	SD	Standard Deviation
I	Sampling Interval	SEM	Structural Equation Modelling
ICR	Inter-Country Coder Reliability Study	SMEG	Subject Matter Expert Group
ICT	Information Communication Technology	SPT	Study Programme Table
IEA	International Association for the Evaluation of Educational Achievement	TA	Test Administrator
INES	OECD Indicators of Education Systems	TAG	Technical Advisory Group
IRT	Item Response Theory	TCS	Target Cluster Size
ISCED	International Standard Classification of Education	TIMSS	Third International Mathematics and Science Study
ISCO	International Standard Classification of Occupations	TIMSS-R	Third International Mathematics and Science Study – Repeat
ISEI	International Socio-Economic Index	VENR	Enrolment for very small schools
MENR	Enrolment for moderately small school	WLE	Weighted Likelihood Estimates
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		



Table of contents

FOREWORD	3
CHAPTER 1 PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT: AN OVERVIEW	19
Participation	21
Features of PISA	22
Managing and implementing PISA	23
Organisation of this report	23
READER'S GUIDE	25
CHAPTER 2 TEST DESIGN AND TEST DEVELOPMENT	27
Test scope and format	28
Test design	28
Test development centres	29
Development timeline	30
The PISA 2006 scientific literacy framework	30
Test development – cognitive items	31
▪ Item development process	31
▪ National item submissions	33
▪ National review of items	34
▪ International item review	35
▪ Preparation of dual (English and French) source versions	35
Test development – attitudinal items	35
Field trial	38
▪ Field trial selection	38
▪ Field trial design	39
▪ Despatch of field trial instruments	40
▪ Field trial coder training	40
▪ Field trial coder queries	40
▪ Field trial outcomes	41
▪ National review of field trial items	42
Main study	42
▪ Main study science items	42
▪ Main study reading items	44
▪ Main study mathematics items	45
▪ Despatch of main study instruments	46
▪ Main study coder training	46
▪ Main study coder query service	46
▪ Review of main study item analyses	47



CHAPTER 3 THE DEVELOPMENT OF THE PISA CONTEXT QUESTIONNAIRES	49
Overview	50
The conceptual structure	51
▪ A conceptual framework for PISA 2006	51
Research areas in PISA 2006	55
The development of the context questionnaires	57
The coverage of the questionnaire material	58
▪ Student questionnaire	58
▪ School questionnaire	59
▪ International options	59
▪ National questionnaire material	60
The implementation of the context questionnaires	60
 CHAPTER 4 SAMPLE DESIGN	 63
Target population and overview of the sampling design	64
Population coverage, and school and student participation rate standards	65
▪ Coverage of the PISA international target population	65
▪ Accuracy and precision	66
▪ School response rates	66
▪ Student response rates	68
Main study school sample	68
▪ Definition of the national target population	68
▪ The sampling frame	69
▪ Stratification	70
▪ Assigning a measure of size to each school	74
▪ School sample selection	74
▪ PISA and TIMSS or PIRLS overlap control	76
▪ Student samples	82
 CHAPTER 5 TRANSLATION AND CULTURAL APPROPRIATENESS OF THE TEST AND SURVEY MATERIAL	 85
Introduction	86
Development of source versions	86
Double translation from two source languages	87
PISA translation and adaptation guidelines	88
Translation training session	89
Testing languages and translation/adaptation procedures	89
International verification of the national versions	91
▪ VegaSuite	93
▪ Documentation	93
▪ Verification of test units	93
▪ Verification of the booklet shell	94
▪ Final optical check	94
▪ Verification of questionnaires and manuals	94
▪ Final check of coding guides	95
▪ Verification outcomes	95



Translation and verification outcomes – national version quality	96
▪ Analyses at the country level.....	96
▪ Analyses at the item level.....	103
▪ Summary of items lost at the national level, due to translation, printing or layout errors.....	104
CHAPTER 6 FIELD OPERATIONS	105
Overview of roles and responsibilities	106
▪ National project managers.....	106
▪ School coordinators.....	107
▪ Test administrators.....	107
▪ School associates.....	108
The selection of the school sample	108
Preparation of test booklets, questionnaires and manuals	108
The selection of the student sample	109
Packaging and shipping materials	110
Receipt of materials at the national centre after testing	110
Coding of the tests and questionnaires	111
▪ Preparing for coding.....	111
▪ Logistics prior to coding.....	113
▪ Single coding design.....	115
▪ Multiple coding.....	117
▪ Managing the process coding.....	118
▪ Cross-national coding.....	120
▪ Questionnaire coding.....	120
Data entry, data checking and file submission	120
▪ Data entry.....	120
▪ Data checking.....	120
▪ Data submission.....	121
▪ After data were submitted.....	121
The main study review	121
CHAPTER 7 QUALITY ASSURANCE	123
PISA quality control	124
▪ Comprehensive operational manuals.....	124
▪ National level implementation planning document.....	124
PISA quality monitoring	124
▪ Field trial and main study review.....	124
▪ Final optical check.....	126
▪ National centre quality monitor (NCQM) visits.....	126
▪ PISA quality monitor (PQM) visits.....	126
▪ Test administration.....	127
▪ Delivery.....	128
CHAPTER 8 SURVEY WEIGHTING AND THE CALCULATION OF SAMPLING VARIANCE	129
Survey weighting	130
The school base weight	131
▪ The school weight trimming factor.....	132

<ul style="list-style-type: none"> ▪ The student base weight 132 ▪ School non-response adjustment..... 132 ▪ Grade non-response adjustment..... 134 ▪ Student non-response adjustment..... 135 ▪ Trimming student weights..... 136 ▪ Comparing the PISA 2006 student non-response adjustment strategy with the strategy used for PISA 2003 136 ▪ The comparison..... 138 	
Calculating sampling variance	139
<ul style="list-style-type: none"> ▪ The balanced repeated replication variance estimator..... 139 ▪ Reflecting weighting adjustments..... 141 ▪ Formation of variance strata..... 141 ▪ Countries where all students were selected for PISA..... 141 	
CHAPTER 9 SCALING PISA COGNITIVE DATA	143
The mixed coefficients multinomial logit model	144
<ul style="list-style-type: none"> ▪ The population model..... 145 ▪ Combined model..... 146 	
Application to PISA	146
<ul style="list-style-type: none"> ▪ National calibrations..... 146 ▪ National reports..... 147 ▪ International calibration 153 ▪ Student score generation..... 153 	
Booklet effects	155
Analysis of data with plausible values	156
Developing common scales for the purposes of trends	157
<ul style="list-style-type: none"> ▪ Linking PISA 2003 and PISA 2006 for reading and mathematics 158 ▪ Uncertainty in the link..... 158 	
CHAPTER 10 DATA MANAGEMENT PROCEDURES	163
Introduction	164
KeyQuest	167
Data management at the national centre	167
<ul style="list-style-type: none"> ▪ National modifications to the database 167 ▪ Student sampling with <i>KeyQuest</i>..... 167 ▪ Data entry quality control 167 	
Data cleaning at ACER	171
<ul style="list-style-type: none"> ▪ Recoding of national adaptations..... 171 ▪ Data cleaning organisation..... 171 ▪ Cleaning reports..... 171 ▪ General recodings..... 171 	
Final review of the data	172
<ul style="list-style-type: none"> ▪ Review of the test and questionnaire data 172 ▪ Review of the sampling data 172 	
Next steps in preparing the international database	172



CHAPTER 11 SAMPLING OUTCOMES	175
Design effects and effective sample sizes	187
▪ Variability of the design effect.....	191
▪ Design effects in PISA for performance variables.....	191
Summary analyses of the design effect	203
▪ Countries with outlying standard errors.....	205
 CHAPTER 12 SCALING OUTCOMES	 207
International characteristics of the item pool	208
▪ Test targeting.....	208
▪ Test reliability.....	208
▪ Domain inter-correlations.....	208
▪ Science scales.....	215
Scaling outcomes	216
▪ National item deletions.....	216
▪ International scaling.....	219
▪ Generating student scale scores.....	219
Test length analysis	219
Booklet effects	221
▪ Overview of the PISA cognitive reporting scales.....	232
▪ PISA overall literacy scales.....	234
▪ PISA literacy scales.....	234
▪ Special purpose scales.....	234
Observations concerning the construction of the PISA overall literacy scales	235
▪ Framework development.....	235
▪ Testing time and item characteristics.....	236
▪ Characteristics of each of the links.....	237
Transforming the plausible values to PISA scales	246
▪ Reading.....	246
▪ Mathematics.....	246
▪ Science.....	246
▪ Attitudinal scales.....	247
Link error	247
 CHAPTER 13 CODING AND MARKER RELIABILITY STUDIES	 249
Homogeneity analyses	251
Multiple marking study outcomes (variance components)	254
▪ Generalisability coefficients.....	254
International coding review	261
▪ Background to changed procedures for PISA 2006.....	261
▪ ICR procedures.....	261
▪ Outcomes.....	264
▪ Cautions.....	270



CHAPTER 14 DATA ADJUDICATION	271
Introduction	272
▪ Implementing the standards – quality assurance	272
▪ Information available for adjudication	273
▪ Data adjudication process	273
General outcomes	274
▪ Overview of response rate issues	274
▪ Detailed country comments	275
CHAPTER 15 PROFICIENCY SCALE CONSTRUCTION	283
Introduction	284
Development of the described scales	285
▪ Stage 1: Identifying possible scales	285
▪ Stage 2: Assigning items to scales	286
▪ Stage 3: Skills audit	286
▪ Stage 4: Analysing field trial data	286
▪ Stage 5: Defining the dimensions	287
▪ Stage 6: Revising and refining with main study data	287
▪ Stage 7: Validating	287
Defining proficiency levels	287
Reporting the results for PISA science	290
▪ Building an item map	290
▪ Levels of scientific literacy	292
▪ Interpreting the scientific literacy levels	299
CHAPTER 16 SCALING PROCEDURES AND CONSTRUCT VALIDATION OF CONTEXT QUESTIONNAIRE DATA	303
Overview	304
Simple questionnaire indices	304
▪ Student questionnaire indices	304
▪ School questionnaire indices	307
▪ Parent questionnaire indices	309
Scaling methodology and construct validation	310
▪ Scaling procedures	310
▪ Construct validation	312
▪ Describing questionnaire scale indices	314
Questionnaire scale indices	315
▪ Student scale indices	315
▪ School questionnaire scale indices	340
▪ Parent questionnaire scale indices	342
▪ The PISA index of economic, social and cultural status (ESCS)	346
CHAPTER 17 VALIDATION OF THE EMBEDDED ATTITUDINAL SCALES	351
Introduction	352
International scalability	353
▪ Analysis of item dimensionality with exploratory and confirmatory factor analysis	353
▪ Fit to item response model	353



▪ Reliability.....	355
▪ Differential item functioning.....	355
▪ Summary of scalability.....	357
Relationship and comparisons with other variables.....	357
▪ Within-country student level correlations with achievement and selected background variables.....	358
▪ Relationships between embedded scales and questionnaire.....	360
▪ Country level correlations with achievement and selected background variables.....	361
▪ Variance decomposition.....	363
▪ Observations from other cross-national data collections.....	363
▪ Summary of relations with other variables.....	364
Conclusion.....	364
CHAPTER 18 INTERNATIONAL DATABASE.....	367
Files in the database.....	368
▪ Student files.....	368
▪ School file.....	370
▪ Parent file.....	370
Records in the database.....	371
▪ Records included in the database.....	371
▪ Records excluded from the database.....	371
Representing missing data.....	371
How are students and schools identified?.....	372
Further information.....	373
REFERENCES.....	375
APPENDICES.....	379
Appendix 1 PISA 2006 main study item pool characteristics.....	380
Appendix 2 Contrast coding used in conditioning.....	389
Appendix 3 Design effect tables.....	399
Appendix 4 Changes to core questionnaire items from 2003 to 2006.....	405
Appendix 5 Mapping of ISCED to years.....	411
Appendix 6 National household possession items.....	412
Appendix 7 Exploratory and confirmatory factor analyses for the embedded items.....	414
Appendix 8 PISA consortium, staff and consultants.....	416



LIST OF BOXES

Box 1.1	Core features of PISA 2006.....	22
---------	---------------------------------	----

LIST OF FIGURES

Figure 2.1	Main study Interest in Science item.....	36
Figure 2.2	Main study Support for Scientific Enquiry item.....	36
Figure 2.3	Field trial Match-the-opinion Responsibility item.....	37
Figure 3.1	Conceptual grid of variable types.....	52
Figure 3.2	The two-dimensional conceptual matrix with examples of variables collected or available from other sources.....	54
Figure 4.1	School response rate standard.....	67
Figure 6.1	Design for the single coding of science and mathematics.....	115
Figure 6.2	Design for the single coding of reading.....	116
Figure 9.1	Example of item statistics in Report 1.....	148
Figure 9.2	Example of item statistics in Report 2.....	149
Figure 9.3	Example of item statistics shown in Graph B.....	150
Figure 9.4	Example of item statistics shown in Graph C.....	151
Figure 9.5	Example of item statistics shown in Table D.....	151
Figure 9.6	Example of summary of dodgy items for a country in Report 3a.....	152
Figure 9.7	Example of summary of dodgy items in Report 3b.....	152
Figure 10.1	Data management in relation to other parts of PISA.....	164
Figure 10.2	Major data management stages in PISA.....	166
Figure 10.3	Validity reports - general hierarchy.....	170
Figure 11.1	Standard error on a mean estimate depending on the intraclass correlation.....	188
Figure 11.2	Relationship between the standard error for the science performance mean and the intraclass correlation within explicit strata (PISA 2006).....	205
Figure 12.1	Item plot for mathematics items.....	210
Figure 12.2	Item plot for reading items.....	211
Figure 12.3	Item plot for science items.....	212
Figure 12.4	Item plot for interest items.....	213
Figure 12.5	Item plot for support items.....	214
Figure 12.6	Scatter plot of per cent correct for reading link items in PISA 2000 and PISA 2003.....	238
Figure 12.7	Scatter plot of per cent correct for reading link items in PISA 2003 and PISA 2006.....	240
Figure 12.8	Scatter plot of per cent correct for mathematics link items in PISA 2003 and PISA 2006.....	242
Figure 12.9	Scatter plot of per cent correct for science link items in PISA 2000 and PISA 2003.....	244
Figure 12.10	Scatter plot of per cent correct for science link items in PISA 2003 and PISA 2006.....	245



Figure 13.1	Variability of the homogeneity indices for science items in field trial	250
Figure 13.2	Average of the homogeneity indices for science items in field trial and main study	251
Figure 13.3	Variability of the homogeneity indices for each science item in the main study	252
Figure 13.4	Variability of the homogeneity indices for each reading item in the main study	252
Figure 13.5	Variability of the homogeneity indices for each mathematics item	252
Figure 13.6	Variability of the homogeneity indices for the participating countries in the main study	253
Figure 13.7	Example of ICR report (reading)	269
<hr/>		
Figure 14.1	Attained school response rates	274
<hr/>		
Figure 15.1	The relationship between items and students on a proficiency scale	285
Figure 15.2	What it means to be at a level	289
Figure 15.3	A map for selected science items	291
Figure 15.4	Summary descriptions of the six proficiency levels on the science scale	294
Figure 15.5	Summary descriptions of six proficiency levels in <i>identifying scientific issues</i>	295
Figure 15.6	Summary descriptions of six proficiency levels in <i>explaining phenomena scientifically</i>	297
Figure 15.7	Summary descriptions of six proficiency levels in <i>using scientific evidence</i>	300
<hr/>		
Figure 16.1	Summed category probabilities for fictitious item	314
Figure 16.2	Fictitious example of an item map	315
Figure 16.3	Scatterplot of country means for ESCS 2003 and ESCS 2006	347
<hr/>		
Figure 17.1	Distribution of item fit mean square statistics for embedded attitude items	354
Figure 17.2	An example of the ESC plot for item S408RNA	356
Figure 17.3	Scatterplot of mean mathematics interest against mean mathematics for PISA 2003	363

LIST OF TABLES

Table 1.1	PISA 2006 participants	21
<hr/>		
Table 2.1	Cluster rotation design used to form test booklets for PISA 2006	29
Table 2.2	Test development timeline for PISA 2006	30
Table 2.3	Science field trial all items	39
Table 2.4	Allocation of item clusters to test booklets for field trial	39
Table 2.5	Science main study items (item format by competency)	43
Table 2.6	Science main study items (item format by knowledge type)	44
Table 2.7	Science main study items (knowledge category by competency)	44
Table 2.8	Reading main study items (item format by aspect)	44
Table 2.9	Reading main study items (item format by text format)	45
Table 2.10	Reading main study items (text type by aspect)	45
Table 2.11	Mathematics main study items (item format by competency cluster)	45
Table 2.12	Mathematics main study items (item format by content category)	46
Table 2.13	Mathematics main study items (content category by competency cluster)	46



Table 3.1	Themes and constructs/variables in PISA 2006.....	56
Table 4.1	Stratification variables	71
Table 4.2	Schedule of school sampling activities	78
Table 5.1	Countries sharing a common version with national adaptations	90
Table 5.2	PISA 2006 translation/adaptation procedures.....	91
Table 5.3	Mean deviation and root mean squared error of the item by country interactions for each version.....	97
Table 5.4	Correlation between national item parameter estimates for Arabic versions.....	99
Table 5.5	Correlation between national item parameter estimates for Chinese versions.....	99
Table 5.6	Correlation between national item parameter estimates for Dutch versions.....	99
Table 5.7	Correlation between national item parameter estimates for English versions.....	99
Table 5.8	Correlation between national item parameter estimates for French versions.....	99
Table 5.9	Correlation between national item parameter estimates for German versions.....	100
Table 5.10	Correlation between national item parameter estimates for Hungarian versions.....	100
Table 5.11	Correlation between national item parameter estimates for Italian versions.....	100
Table 5.12	Correlation between national item parameter estimates for Portuguese versions.....	100
Table 5.13	Correlation between national item parameter estimates for Russian versions.....	100
Table 5.14	Correlation between national item parameter estimates for Spanish versions	100
Table 5.15	Correlation between national item parameter estimates for Swedish versions	100
Table 5.16	Correlation between national item parameter estimates within countries.....	101
Table 5.17	Variance estimate.....	102
Table 5.18	Variance estimates	103
Table 6.1	Design for the multiple coding of science and mathematics.....	118
Table 6.2	Design for the multiple coding of reading.....	118
Table 8.1	Non-response classes	133
Table 9.1	Deviation contrast coding scheme	154
Table 10.1	Double entry discrepancies per country: field trial data.....	169
Table 11.1	Sampling and coverage rates.....	178
Table 11.2	School response rates before replacement.....	182
Table 11.3	School response rates after replacement.....	184
Table 11.4	Student response rates after replacement.....	185
Table 11.5	Standard errors for the PISA 2006 combined science scale	189
Table 11.6	Design effect 1 by country, by domain and cycle.....	193
Table 11.7	Effective sample size 1 by country, by domain and cycle.....	194
Table 11.8	Design effect 2 by country, by domain and cycle.....	195
Table 11.9	Effective sample size 2 by country, by domain and cycle.....	196
Table 11.10	Design effect 3 by country, by domain and by cycle.....	197



Table 11.11	Effective sample size 3 by country, by domain and cycle	198
Table 11.12	Design effect 4 by country, by domain and cycle.....	199
Table 11.13	Effective sample size 4 by country, by domain and cycle	200
Table 11.14	Design effect 5 by country, by domain and cycle.....	201
Table 11.15	Effective sample size 5 by country, by domain and cycle	202
Table 11.16	Median of the design effect 3 per cycle and per domain across the 35 countries that participated in every cycle.....	203
Table 11.17	Median of the standard errors of the student performance mean estimate for each domain and PISA cycle for the 35 countries that participated in every cycle	203
Table 11.18	Median of the number of participating schools for each domain and PISA cycle for the 35 countries that participated in every cycle.....	204
Table 11.19	Median of the school variance estimate for each domain and PISA cycle for the 35 countries that participated in every cycle.....	204
Table 11.20	Median of the intraclass correlation for each domain and PISA cycle for the 35 countries that participated in every cycle.....	204
Table 11.21	Median of the within explicit strata intraclass correlation for each domain and PISA cycle for the 35 countries that participated in every cycle	205
Table 11.22	Median of the percentages of school variances explained by explicit stratification variables, for each domain and PISA cycle for the 35 countries that participated in every cycle	205
<hr/>		
Table 12.1	Number of sampled student by country and booklet.....	209
Table 12.2	Reliabilities of each of the four overall scales when scaled separately.....	215
Table 12.3	Latent correlation between the five domains	215
Table 12.4	Latent correlation between science scales	215
Table 12.5	Items deleted at the national level	216
Table 12.6	Final reliability of the PISA scales	216
Table 12.7	National reliabilities for the main domains.....	217
Table 12.8	National reliabilities for the science subscales.....	218
Table 12.9	Average number of not-reached items and missing items by booklet.....	219
Table 12.10	Average number of not-reached items and missing items by country.....	220
Table 12.11	Distribution of not-reached items by booklet	221
Table 12.12	Estimated booklet effects on the PISA scale.....	221
Table 12.13	Estimated booklet effects in logits	221
Table 12.14	Variance in mathematics booklet means	222
Table 12.15	Variance in reading booklet means.....	224
Table 12.16	Variance in science booklet means.....	226
Table 12.17	Variance in interest booklet means	228
Table 12.18	Variance in support booklet means.....	230
Table 12.19	Summary of PISA cognitive reporting scales	233
Table 12.20	Linkage types among PISA domains 2000-2006	235
Table 12.21	Number of unique item minutes for each domain for each PISA assessments.....	237
Table 12.22	Numbers of link items between successive PISA assessments.....	237
Table 12.23	Per cent correct for reading link items in PISA 2000 and PISA 2003	238
Table 12.24	Per cent correct for reading link items in PISA 2003 and PISA 2006	239
Table 12.25	Per cent correct for mathematics link items in PISA 2003 and PISA 2006	241

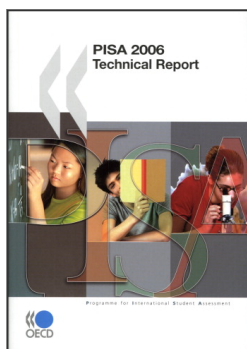
Table 12.26	Per cent correct for science link items in PISA 2000 and PISA 2003	243
Table 12.27	Per cent correct for science link items in PISA 2003 and PISA 2006	245
Table 12.28	Link error estimates	247
<hr/>		
Table 13.1	Variance components for mathematics.....	255
Table 13.2	Variance components for science	256
Table 13.3	Variance components for reading.....	257
Table 13.4	Generalisability estimates for mathematics.....	258
Table 13.5	Generalisability estimates for science	259
Table 13.6	Generalisability estimates for reading	260
Table 13.7	Examples of flagged cases	263
Table 13.8	Count of analysis groups showing potential bias, by domain.....	264
Table 13.9	Comparison of codes assigned by verifier and adjudicator	265
Table 13.10	Outcomes of ICR analysis part 1	265
Table 13.11	ICR outcomes by country and domain	266
<hr/>		
Table 15.1	Scientific literacy performance band definitions on the PISA scale	293
<hr/>		
Table 16.1	ISCO major group white-collar/blue-collar classification	306
Table 16.2	ISCO occupation categories classified as science-related occupations	307
Table 16.3	OECD means and standard deviations of WL estimates	311
Table 16.4	Median, minimum and maximum percentages of between-school variance for student-level indices across countries.....	313
Table 16.5	Household possessions and home background indices.....	316
Table 16.6	Scale reliabilities for home possession indices in OECD countries	317
Table 16.7	Scale reliabilities for home possession indices in partner countries/economies	318
Table 16.8	Item parameters for interest in science learning (INTSCIE).....	318
Table 16.9	Item parameters for enjoyment of science (JOYSCIE)	319
Table 16.10	Model fit and estimated latent correlations for interest in and enjoyment of science learning.....	319
Table 16.11	Scale reliabilities for interest in and enjoyment of science learning.....	320
Table 16.12	Item parameters for instrumental motivation to learn science (INSTSCIE).....	320
Table 16.13	Item parameters for future-oriented science motivation (SCIEFUT).....	321
Table 16.14	Model fit and estimated latent correlations for motivation to learn science	321
Table 16.15	Scale reliabilities for instrumental and future-oriented science motivation.....	322
Table 16.16	Item parameters for science self-efficacy (SCIEEFF).....	322
Table 16.17	Item parameters for science self-concept (SCSCIE).....	323
Table 16.18	Model fit and estimated latent correlations for science self-efficacy and science self-concept.....	323
Table 16.19	Scale reliabilities for science self-efficacy and science self-concept.....	324
Table 16.20	Item parameters for general value of science (GENSCIE).....	324
Table 16.21	Item parameters for personal value of science (PERSCIE).....	325
Table 16.22	Model fit and estimated latent correlations for general and personal value of science.....	325
Table 16.23	Scale reliabilities for general and personal value of science.....	326
Table 16.24	Item parameters for science activities (SCIEACT)	326



Table 16.25	Scale reliabilities for the science activities index	327
Table 16.26	Item parameters for awareness of environmental issues (ENVAWARE)	327
Table 16.27	Item parameters for perception of environmental issues (ENVPERC)	328
Table 16.28	Item parameters for environmental optimism (ENVOPT)	328
Table 16.29	Item parameters for responsibility for sustainable development (RESPDEV)	328
Table 16.30	Model fit environment-related constructs	329
Table 16.31	Estimated latent correlations for environment-related constructs	329
Table 16.32	Scale reliabilities for environment-related scales in OECD countries	330
Table 16.33	Scale reliabilities for environment-related scales in non-OECD countries	330
Table 16.34	Item parameters for school preparation for science career (CARPREP)	331
Table 16.35	Item parameters for student information on science careers (CARINFO)	331
Table 16.36	Model fit and estimated latent correlations for science career preparation indices	332
Table 16.37	Scale reliabilities for science career preparation indices	332
Table 16.38	Item parameters for science teaching: interaction (SCINTACT)	333
Table 16.39	Item parameters for science teaching: hands-on activities (SCHANDS)	333
Table 16.40	Item parameters for science teaching: student investigations (SCINVEST)	333
Table 16.41	Item parameters for science teaching: focus on models or applications (SCAPPLY)	334
Table 16.42	Model fit for CFA with science teaching and learning	334
Table 16.43	Estimated latent correlations for constructs related to science teaching and learning	335
Table 16.44	Scale reliabilities for scales to science teaching and learning in OECD countries	336
Table 16.45	Scale reliabilities for scales to science teaching and learning in partner countries/economies	336
Table 16.46	Item parameters for ICT Internet/entertainment use (INTUSE)	337
Table 16.47	Item parameters for ICT program/software use (PRGUSE)	337
Table 16.48	Item parameters for ICT self-confidence in Internet tasks (INTCONF)	337
Table 16.49	Item parameters for ICT self-confidence in high-level ICT tasks (HIGHCONF)	338
Table 16.50	Model fit for CFA with ICT familiarity items	338
Table 16.51	Estimated latent correlations for constructs related to ICT familiarity	339
Table 16.52	Scale reliabilities for ICT familiarity scales	339
Table 16.53	Item parameters for teacher shortage (TCSHORT)	340
Table 16.54	Item parameters for quality of educational resources (SCMATEDU)	340
Table 16.55	Item parameters for school activities to promote the learning of science (SCIPROM)	341
Table 16.56	Item parameters for school activities for learning environmental topics (ENVLEARN)	341
Table 16.57	Scale reliabilities for school-level scales in OECD countries	341
Table 16.58	Scale reliabilities for environment-related scales in partner countries/economies	342
Table 16.59	Item parameters for science activities at age 10 (PQSCIACT)	343
Table 16.60	Item parameters for parent's perception of school quality (PQSCHOOL)	343
Table 16.61	Item parameters for parent's views on importance of science (PQSCIMP)	343
Table 16.62	Item parameters for parent's reports on science career motivation (PQSCCAR)	344
Table 16.63	Item parameters for parent's view on general value of science (PQGENSCI)	344
Table 16.64	Item parameters for parent's view on personal value of science (PQPERSCI)	344
Table 16.65	Item parameters for parent's perception of environmental issues (PQENPERC)	345
Table 16.66	Item parameters for parent's environmental optimism (PQENVOPT)	345



Table 16.67	Scale reliabilities for parent questionnaire scales.....	345
Table 16.68	Factor loadings and internal consistency of ESCS 2006 in OECD countries.....	347
Table 16.69	Factor loadings and internal consistency of ESCS 2006 in partner countries/economies.....	348
<hr/>		
Table 17.1	Student-level latent correlations between mathematics, reading, science, embedded interest and embedded support.....	354
Table 17.2	Summary of the IRT scaling results across countries.....	355
Table 17.3	Gender DIF table for embedded attitude items.....	357
Table 17.4	Correlation amongst attitudinal scales, performance scales and HISEI.....	358
Table 17.5	Correlations for science scale.....	359
Table 17.6	Loadings of the achievement, interest and support variables on three varimax rotated components.....	360
Table 17.7	Correlation between embedded attitude scales and questionnaire attitude scales.....	361
Table 17.8	Rank order correlation five test domains, questionnaire attitude scales and HISEI.....	362
Table 17.9	Intra-class correlation (rho).....	362
<hr/>		
Table A1.1	2006 Main study reading item classification.....	380
Table A1.2	2006 Main study mathematics item classification.....	381
Table A1.3	2006 Main study science item classification (cognitive).....	383
Table A1.4	2006 Main study science embedded item classification (interest in learning science topics).....	387
Table A1.5	2006 Main study science embedded item classification (support for scientific enquiry).....	388
<hr/>		
Table A2.1	2006 Main study contrast coding used in conditioning for the student questionnaire variables.....	389
Table A2.2	2006 Main study contrast coding used in conditioning for the ICT questionnaire variables.....	396
Table A2.3	2006 Main study contrast coding used in conditioning for the parent questionnaire variables and other variables.....	397
<hr/>		
Table A3.1	Standard errors of the student performance mean estimate by country, by domain and cycle.....	399
Table A3.2	Sample sizes by country and cycle.....	400
Table A3.3	School variance estimate by country, by domain and cycle.....	401
Table A3.4	Intraclass correlation by country, by domain and cycle.....	402
Table A3.5	Within explicit strata intraclass correlation by country, by domain and cycle.....	403
Table A3.6	Percentages of school variance explained by explicit stratification variables, by domain and cycle.....	404
<hr/>		
Table A4.1	Student questionnaire.....	405
Table A4.2	ICT familiarity questionnaire.....	407
Table A4.3	School questionnaire.....	408
<hr/>		
Table A5.1	Mapping of ISCED to accumulated years of education.....	411
<hr/>		
Table A6.1	National household possession items.....	412
<hr/>		
Table A7.1	Exploratory and confirmatory factor analyses (EFA and CFA) for the embedded items.....	414



From:
PISA 2006 Technical Report

Access the complete publication at:
<https://doi.org/10.1787/9789264048096-en>

Please cite this chapter as:

OECD (2009), "Sample design", in *PISA 2006 Technical Report*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264048096-5-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.