



Annex A

PISA 2012 TECHNICAL BACKGROUND

All figures and tables in Annex A are available on line

Annex A1: Indices from the student context questionnaires

Annex A2: The PISA target population, the PISA samples and the definition of schools
<http://dx.doi.org/10.1787/888933003725>

Annex A3: Technical notes on analyses in this volume

Annex A4: Quality assurance

Annex A5: The problem-solving assessment design

Annex A6: Technical note on Brazil
<http://dx.doi.org/10.1787/888933003744>

Notes regarding Cyprus

Note by Turkey: The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

A note regarding Israel

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.



ANNEX A1

INDICES FROM THE STUDENT CONTEXT QUESTIONNAIRES

Explanation of the indices

This section explains the indices derived from the student context questionnaires used in PISA 2012.

Several PISA measures reflect indices that summarise responses from students, their parents or school representatives (typically principals) to a series of related questions. The questions were selected from a larger pool of questions on the basis of theoretical considerations and previous research. The *PISA 2012 Assessment and Analytical Framework* (OECD, 2013a) provides an in-depth description of this conceptual framework. Structural equation modelling was used to confirm the theoretically expected behaviour of the indices and to validate their comparability across countries. For this purpose, a model was estimated separately for each country and collectively for all OECD countries. For a detailed description of other PISA indices and details on the methods, see the *PISA 2012 Technical Report* (OECD, forthcoming).

There are two types of indices: simple indices and scale indices.

Simple indices are the variables that are constructed through the arithmetic transformation or recoding of one or more items, in exactly the same way across assessments. Here, item responses are used to calculate meaningful variables, such as the recoding of the four-digit ISCO-08 codes into “Highest parents’ socio-economic index (HISEI)” or, teacher-student ratio based on information from the school questionnaire.

Scale indices are the variables constructed through the scaling of multiple items. Unless otherwise indicated, the index was scaled using a weighted likelihood estimate (WLE) (Warm, 1989), using a one-parameter item response model (a partial credit model was used in the case of items with more than two categories). For details on how each scale index was constructed see the *PISA 2012 Technical Report* (OECD, forthcoming). In general, the scaling was done in three stages:

- The item parameters were estimated from equal-sized subsamples of students from all participating countries and economies.
- The estimates were computed for all students and all schools by anchoring the item parameters obtained in the preceding step.
- The indices were then standardised so that the mean of the index value for the OECD student population was zero and the standard deviation was one (countries being given equal weight in the standardisation process).

Sequential codes were assigned to the different response categories of the questions in the sequence in which the latter appeared in the student, school or parent questionnaires. Where indicated in this section, these codes were inverted for the purpose of constructing indices or scales. Negative values for an index do not necessarily imply that students responded negatively to the underlying questions. A negative value merely indicates that the respondents answered less positively than all respondents did on average across OECD countries. Likewise, a positive value on an index indicates that the respondents answered more favourably, or more positively, than respondents did, on average, across OECD countries. Terms enclosed in brackets < > in the following descriptions were replaced in the national versions of the student, school and parent questionnaires by the appropriate national equivalent. For example, the term <qualification at ISCED level 5A> was translated in the United States into “Bachelor’s degree, post-graduate certificate program, Master’s degree program or first professional degree program”. Similarly the term <classes in the language of assessment> in Luxembourg was translated into “German classes” or “French classes” depending on whether students received the German or French version of the assessment instruments.

In addition to simple and scaled indices described in this annex, there are a number of variables from the questionnaires that correspond to single items not used to construct indices. These non-recoded variables have prefix of “ST” for the questionnaire items in the student background questionnaire, and “IC” for the items in the information and communication technology familiarity questionnaire. All the context questionnaires as well as the PISA international database, including all variables, are available through www.pisa.oecd.org.

Student-level simple indices

Study programme

In PISA 2012, study programmes available to 15-year-old students in each country were collected both through the student tracking form and the student questionnaire. All study programmes were classified using ISCED (OECD, 1999). In the PISA international database, all national programmes are indicated in a variable (PROGN) where the first six digits refer to the national centre code and the last two digits to the national study programme code.

The following internationally comparable indices were derived from the data on study programmes:

- Programme level (ISCEDL) indicates whether students are (1) primary education level (ISCED 1); (2) lower-secondary education level (ISCED 2); or (3) upper secondary education level (ISCED 3).
- Programme designation (ISCEDD) indicates the designation of the study programme: (1) = “A” (general programmes designed to give access to the next programme level); (2) = “B” (programmes designed to give access to vocational studies at the next programme level); (3) = “C” (programmes designed to give direct access to the labour market); or (4) = “M” (modular programmes that combine any or all of these characteristics).



- Programme orientation (ISCEDO) indicates whether the programme's curricular content is (1) general; (2) pre-vocational; (3) vocational; or (4) modular programmes that combine any or all of these characteristics.

Occupational status of parents

Occupational data for both a student's father and a student's mother were obtained by asking open-ended questions in the student questionnaire. The responses were coded to four-digit ISCO codes (ILO, 1990) and then mapped to the SEI index of Ganzeboom et al. (1992). Higher scores of SEI indicate higher levels of occupational status. The following three indices are obtained:

- Mother's occupational status (OCOD1).
- Father's occupational status (OCOD2).
- The highest occupational level of parents (HISEI) corresponds to the higher SEI score of either parent or to the only available parent's SEI score.

Some of the analyses distinguish between four different categories of occupations by the major groups identified by the ISCO coding of the highest parental occupation: Elementary (ISCO 9), semi-skilled blue-collar (ISCO 6, 7 and 8), semi-skilled white-collar (ISCO 4 and 5), skilled (ISCO 1, 2 and 3). This classification follows the same methodology used in other OECD publications such as *Education at a Glance* (OECD, 2013b) and the *OECD Skills Outlook* (OECD, 2013c).¹

Education level of parents

The education level of parents is classified using ISCED (OECD, 1999) based on students' responses in the student questionnaire.

As in PISA 2000, 2003, 2006 and 2009, indices were constructed by selecting the highest level for each parent and then assigning them to the following categories: (0) None, (1) ISCED 1 (primary education), (2) ISCED 2 (lower secondary), (3) ISCED 3B or 3C (vocational/pre-vocational upper secondary), (4) ISCED 3A (upper secondary) and/or ISCED 4 (non-tertiary post-secondary), (5) ISCED 5B (vocational tertiary), (6) ISCED 5A, 6 (theoretically oriented tertiary and post-graduate). The following three indices with these categories are developed:

- Mother's education level (MISCED).
- Father's education level (FISCED).
- Highest education level of parents (HISCED) corresponds to the higher ISCED level of either parent.

Highest education level of parents was also converted into the number of years of schooling (PARED). For the conversion of level of education into years of schooling, see Table A1.1 in Volume I (OECD, 2013d).

Immigration background

Information on the country of birth of students and their parents is collected in a similar manner as in PISA 2000, PISA 2003 and PISA 2006 by using nationally specific ISO coded variables. The ISO codes of the country of birth for students and their parents are available in the PISA international database (COBN_S, COBN_M, and COBN_F).

The index on immigrant background (IMMIG) has the following categories: (1) native students (those students born in the country of assessment, or those with at least one parent born in that country; students who were born abroad with at least one parent born in the country of assessment are also classified as native students), (2) second-generation students (those born in the country of assessment but whose parents were born in another country) and (3) first-generation students (those born outside the country of assessment and whose parents were also born in another country). Students with missing responses for either the student or for both parents, or for all three questions have been given missing values for this variable.

Use of computers at home

An indicator about students' use of desktop, laptop or tablet computers at home was derived using their responses to the questionnaire on students' familiarity with information and communication. Three items in question IC01 ("Are any of these devices available for you to use at home?") were used: Desktop computer; Portable laptop or notebook; <Tablet computer> (e.g. <iPad®>, <BlackBerry® PlayBook™>). Students who answered "Yes, and I use it" to at least one of these questions have a value of 1 for this indicator.

Use of computers at school

An indicator about students' use of desktop, laptop or tablet computers at school was derived using their responses to the questionnaire on students' familiarity with information and communication technology (ICT). Three items in question IC02 ("Are any of these devices available for you to use at school?") were used: Desktop computer; Portable laptop or notebook; <Tablet computer> (e.g. <iPad®>, <BlackBerry® PlayBook™>). Students who answered "Yes, and I use it" to at least one of these questions have a value of 1 for this indicator.

1. Note that for ISCO coding 0 "Arm forces", the following recoding was followed: "Officers" were coded as "Managers" (ISCO 1), and "Other armed forces occupations" (drivers, gunners, seaman, generic armed forces) as "Plant and Machine operators" (ISCO 8). In addition, all answers starting with "97" (housewives, students, and "vague occupations") were coded into missing.



Student-level scale indices

In order to obtain trends for socio-economic scale indices from 2000 to 2012, the scaling of the indices WEALTH, HEDRES, CULTPOSS, HOMEPOS and ESCS was based on data from all cycles from 2000 to 2012.

Family wealth

The *index of family wealth* (WEALTH) is based on students' responses on whether they had the following at home: a room of their own, a link to the Internet, a dishwasher (treated as a country-specific item), a DVD player, and three other country-specific items; and their responses on the number of cellular phones, televisions, computers, cars and the number of rooms with a bath or shower.

Home educational resources

The *index of home educational resources* (HEDRES) is based on the items measuring the existence of educational resources at home including a desk and a quiet place to study, a computer that students can use for schoolwork, educational software, books to help with students' school work, technical reference books and a dictionary.

Cultural possessions

The *index of cultural possessions* (CULTPOSS) is based on the students' responses to whether they had the following at home: classic literature, books of poetry and works of art.

Economic, social and cultural status

The *PISA index of economic, social and cultural status* (ESCS) was derived from the following three indices: *highest occupational status of parents* (HISEI), *highest education level of parents* in years of education according to ISCED (PARED), and *home possessions* (HOMEPOS). The *index of home possessions* (HOMEPOS) comprises all items on the indices of WEALTH, CULTPOSS and HEDRES, as well as books in the home recoded into a four-level categorical variable (0-10 books, 11-25 or 26-100 books, 101-200 or 201-500 books, more than 500 books).

The *PISA index of economic, social and cultural status* (ESCS) was derived from a principal component analysis of standardised variables (each variable has an OECD mean of zero and a standard deviation of one), taking the factor scores for the first principal component as measures of the *PISA index of economic, social and cultural status*.

Principal component analysis was also performed for each participating country to determine to what extent the components of the index operate in similar ways across countries. The analysis revealed that patterns of factor loading were very similar across countries, with all three components contributing to a similar extent to the index (for details on reliability and factor loadings, see the *PISA 2012 Technical Report* (OECD, forthcoming).

The imputation of components for students with missing data on one component was done on the basis of a regression on the other two variables, with an additional random error component. The final values on the *PISA index of economic, social and cultural status* (ESCS) for PISA 2012 have an OECD mean of zero and a standard deviation of one.

Perseverance

The *index of perseverance* (PERSEV) was constructed using student responses (ST93) over whether they report that the following statements describe them very much, mostly, somewhat, not much, not at all: When confronted with a problem, I give up easily; I put off difficult problems; I remain interested in the tasks that I start; I continue working on tasks until everything is perfect; When confronted with a problem, I do more than what is expected of me.

Openness to problem solving

The *index of openness to problem solving* (OPENPS) was constructed using student responses (ST94) over whether they report that the following statements describe them very much, mostly, somewhat, not much, not at all: I can handle a lot of information; I am quick to understand things; I seek explanations of things; I can easily link facts together; I like to solve complex problems.

The rotated design of the student questionnaire

A major innovation in PISA 2012 is the rotated design of the student questionnaire. One of the main reasons for a rotated design, which has previously been implemented for the cognitive assessment, was to extend the content coverage of the student questionnaire. Table A1.1 provides an overview of the rotation design and content of questionnaire forms for the main survey.

The *PISA 2012 Technical Report* (OECD, forthcoming) provides all details regarding the rotated design of the student questionnaire in PISA 2012, including its implications in terms of (a) proficiency estimates, (b) international reports and trends, (c) further analyses, (d) structure and documentation of the international database, and (e) logistics have been discussed elsewhere. The rotated design has negligible implications for proficiency estimates and correlations of proficiency estimates with context constructs. The international database (available at www.pisa.oecd.org) includes all background variables for each student. The variables based on questions that students answered reflect their responses; those that are based on questions that were not administered show a distinctive missing code. Rotation allows the estimation of a full co-variance matrix which means that all variables can be correlated with all other variables. It does not affect conclusions in terms of whether or not an effect would be considered significant in multilevel models.

**Table A1.1 Student questionnaire rotation design**

Form A	Common Question Set (all forms)	Question Set 1 – Mathematics Attitudes / Problem Solving	Question Set 3 – Opportunity to Learn / Learning Strategies
Form B	Common Question Set (all forms)	Question Set 2 – School Climate / Attitudes towards School / Anxiety	Question Set 1 – Mathematics Attitudes / Problem Solving
Form C	Common Question Set (all forms)	Question Set 3 – Opportunity to Learn / Learning Strategies	Question Set 2 – School Climate / Attitudes towards School / Anxiety

Note: For details regarding the questions in each question set, please refer to the *PISA 2012 Technical Report* (OECD, forthcoming).

References

Ganzeboom, H.B.G., P. De Graaf, and D.J. Treiman (with J. De Leeuw) (1992), "A Standard International Socio-Economic Index of Occupational Status", *Social Science Research* (21-1), pp. 1-56.

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

OECD (forthcoming), *PISA 2012 Technical Report*, PISA, OECD Publishing.

OECD (2013a), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, PISA, OECD Publishing.

<http://dx.doi.org/10.1787/9789264190511-en>

OECD (2013b), *Education at a Glance 2013: OECD Indicators*, OECD Publishing.

<http://dx.doi.org/10.1787/eag-2013-en>

OECD (2013c), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing.

<http://dx.doi.org/10.1787/9789264204256-en>

OECD (2013d), *PISA 2012 Results: What Students Know and Can Do: Student Performance in Mathematics, Reading and Science (Volume I)*, PISA, OECD Publishing.

<http://dx.doi.org/10.1787/9789264201118-en>

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

<http://dx.doi.org/10.1787/9789264006416-en>

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

<http://www.oecd.org/education/skills-beyond-school/1962350.pdf>

Warm, T.A. (1989), "Weighted likelihood estimation of ability in item response theory", *Psychometrika*, Volume 54, Issue 3, pp 427-450.

<http://dx.doi.org/10.1007/BF02294627>



ANNEX A2

THE PISA TARGET POPULATION, THE PISA SAMPLES AND THE DEFINITION OF SCHOOLS

Definition of the PISA target population

PISA 2012 provides an assessment of the cumulative yield of education and learning at a point at which most young adults are still enrolled in initial education.

A major challenge for an international survey is to ensure that international comparability of national target populations is guaranteed in such a venture.

Differences between countries in the nature and extent of pre-primary education and care, the age of entry into formal schooling and the institutional structure of education systems do not allow the definition of internationally comparable grade levels of schooling. Consequently, international comparisons of education performance typically define their populations with reference to a target age group. Some previous international assessments have defined their target population on the basis of the grade level that provides maximum coverage of a particular age cohort. A disadvantage of this approach is that slight variations in the age distribution of students across grade levels often lead to the selection of different target grades in different countries, or between education systems within countries, raising serious questions about the comparability of results across, and at times within, countries. In addition, because not all students of the desired age are usually represented in grade-based samples, there may be a more serious potential bias in the results if the unrepresented students are typically enrolled in the next higher grade in some countries and the next lower grade in others. This would exclude students with potentially higher levels of performance in the former countries and students with potentially lower levels of performance in the latter.

In order to address this problem, PISA uses an age-based definition for its target population, i.e. a definition that is not tied to the institutional structures of national education systems. PISA assesses students who were aged between 15 years and 3 (complete) months and 16 years and 2 (complete) months at the beginning of the assessment period, plus or minus a 1 month allowable variation, and who were enrolled in an educational institution with Grade 7 or higher, regardless of the grade levels or type of institution in which they were enrolled, and regardless of whether they were in full-time or part-time education. Educational institutions are generally referred to as schools in this publication, although some educational institutions (in particular, some types of vocational education establishments) may not be termed schools in certain countries. As expected from this definition, the average age of students across OECD countries was 15 years and 9 months. The range in country means was 2 months and 5 days (0.18 years), from the minimum country mean of 15 years and 8 months to the maximum country mean of 15 years and 10 months.

Given this definition of population, PISA makes statements about the knowledge and skills of a group of individuals who were born within a comparable reference period, but who may have undergone different educational experiences both in and outside of schools. In PISA, these knowledge and skills are referred to as the yield of education at an age that is common across countries. Depending on countries' policies on school entry, selection and promotion, these students may be distributed over a narrower or a wider range of grades across different education systems, tracks or streams. It is important to consider these differences when comparing PISA results across countries, as observed differences between students at age 15 may no longer appear as students' educational experiences converge later on.

If a country's scale scores in reading, scientific or mathematical literacy are significantly higher than those in another country, it cannot automatically be inferred that the schools or particular parts of the education system in the first country are more effective than those in the second. However, one can legitimately conclude that the cumulative impact of learning experiences in the first country, starting in early childhood and up to the age of 15, and embracing experiences both in school, home and beyond, have resulted in higher outcomes in the literacy domains that PISA measures.

The PISA target population did not include residents attending schools in a foreign country. It does, however, include foreign nationals attending schools in the country of assessment.

To accommodate countries that desired grade-based results for the purpose of national analyses, PISA 2012 provided a sampling option to supplement age-based sampling with grade-based sampling.

Population coverage

All countries attempted to maximise the coverage of 15-year-olds enrolled in education in their national samples, including students enrolled in special educational institutions. As a result, PISA 2012 reached standards of population coverage that are unprecedented in international surveys of this kind.

The sampling standards used in PISA permitted countries to exclude up to a total of 5% of the relevant population either by excluding schools or by excluding students within schools. All but eight countries, Luxembourg (8.40%), Canada (6.38%), Denmark (6.18%), Norway (6.11%), Estonia (5.80%), Sweden (5.44%), the United Kingdom (5.43%) and the United States (5.35%), achieved this standard, and in 30 countries and economies, the overall exclusion rate was less than 2%. When language exclusions were accounted for (i.e. removed from the overall exclusion rate), Norway, Sweden, the United Kingdom and the United States no longer had an exclusion rate greater than 5%. For details, see www.pisa.oecd.org.



Exclusions within the above limits include:

- At the school level: *i*) schools that were geographically inaccessible or where the administration of the PISA assessment was not considered feasible; and *ii*) schools that provided teaching only for students in the categories defined under “within-school exclusions”, such as schools for the blind. The percentage of 15-year-olds enrolled in such schools had to be less than 2.5% of the nationally desired target population [0.5% maximum for *i*) and 2% maximum for *ii*)]. The magnitude, nature and justification of school-level exclusions are documented in the *PISA 2012 Technical Report* (OECD, forthcoming).
- At the student level: *i*) students with an intellectual disability; *ii*) students with a functional disability; *iii*) students with limited assessment language proficiency; *iv*) other – a category defined by the national centres and approved by the international centre; and *v*) students taught in a language of instruction for the main domain for which no materials were available. Students could not be excluded solely because of low proficiency or common discipline problems. The percentage of 15-year-olds excluded within schools had to be less than 2.5% of the nationally desired target population.

Table A2.1 describes the target population of the countries participating in PISA 2012. Further information on the target population and the implementation of PISA sampling standards can be found in the *PISA 2012 Technical Report* (OECD, forthcoming).

- **Column 1** shows the *total number of 15-year-olds* according to the most recent available information, which in most countries meant the year 2011 as the year before the assessment.
- **Column 2** shows the number of 15-year-olds enrolled in schools in Grade 7 or above (as defined above), which is referred to as the *eligible population*.
- **Column 3** shows the *national desired target population*. Countries were allowed to exclude up to 0.5% of students a priori from the eligible population, essentially for practical reasons. The following a priori exclusions exceed this limit but were agreed with the PISA Consortium: Belgium excluded 0.23% of its population for a particular type of student educated while working; Canada excluded 1.14% of its population from Territories and Aboriginal reserves; Chile excluded 0.04% of its students who live in Easter Island, Juan Fernandez Archipelago and Antarctica; Indonesia excluded 1.55% of its students from two provinces because of operational reasons; Ireland excluded 0.05% of its students in three island schools off the west coast; Latvia excluded 0.08% of its students in distance learning schools; and Serbia excluded 2.11% of its students taught in Serbian in Kosovo.
- **Column 4** shows the *number of students enrolled in schools that were excluded from the national desired target population* either from the sampling frame or later in the field during data collection.
- **Column 5** shows the *size of the national desired target population after subtracting the students enrolled in excluded schools*. This is obtained by subtracting Column 4 from Column 3.
- **Column 6** shows the *percentage of students enrolled in excluded schools*. This is obtained by dividing Column 4 by Column 3 and multiplying by 100.
- **Column 7** shows the *number of students participating in PISA 2012*. Note that in some cases this number does not account for 15-year-olds assessed as part of additional national options.
- **Column 8** shows the *weighted number of participating students*, i.e. the number of students in the nationally defined target population that the PISA sample represents.
- Each country attempted to maximise the coverage of the PISA target population within the sampled schools. In the case of each sampled school, all eligible students, namely those 15 years of age, regardless of grade, were first listed. Sampled students who were to be excluded had still to be included in the sampling documentation, and a list drawn up stating the reason for their exclusion. **Column 9** indicates the *total number of excluded students*, which is further described and classified into specific categories in Table A2.2.
- **Column 10** indicates the *weighted number of excluded students*, i.e. the overall number of students in the nationally defined target population represented by the number of students excluded from the sample, which is also described and classified by exclusion categories in Table A2.2. Excluded students were excluded based on five categories: *i*) students with an intellectual disability – the student has a mental or emotional disability and is cognitively delayed such that he/she cannot perform in the PISA testing situation; *ii*) students with a functional disability – the student has a moderate to severe permanent physical disability such that he/she cannot perform in the PISA testing situation; *iii*) students with a limited assessment language proficiency – the student is unable to read or speak any of the languages of the assessment in the country and would be unable to overcome the language barrier in the testing situation (typically a student who has received less than one year of instruction in the languages of the assessment may be excluded); *iv*) other – a category defined by the national centres and approved by the international centre; and *v*) students taught in a language of instruction for the main domain for which no materials were available.
- **Column 11** shows the *percentage of students excluded within schools*. This is calculated as the weighted number of excluded students (Column 10), divided by the weighted number of excluded and participating students (Column 8 plus Column 10), then multiplied by 100.
- **Column 12** shows the *overall exclusion rate*, which represents the weighted percentage of the national desired target population excluded from PISA either through school-level exclusions or through the exclusion of students within schools. It is calculated as the school-level exclusion rate (Column 6 divided by 100) plus within-school exclusion rate (Column 11 divided by 100) multiplied by 1 minus the school-level exclusion rate (Column 6 divided by 100). This result is then multiplied by 100. Eight countries, Canada, Denmark, Estonia, Luxembourg, Norway, Sweden, the United Kingdom and the United States, had exclusion rates higher than 5%. When language exclusions were accounted for (i.e. removed from the overall exclusion rate), Norway, Sweden, the United Kingdom and the United States no longer had an exclusion rate greater than 5%.

[Part 1/2]
Table A2.1 PISA target populations and samples

		Population and sample information							
		Total population of 15-year-olds	Total enrolled population of 15-year-olds at Grade 7 or above	Total in national desired target population	Total school-level exclusions	Total in national desired target population after all school exclusions and before within-school exclusions	School-level exclusion rate (%)	Number of participating students	Weighted number of participating students
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
OECD	Australia	291 967	288 159	288 159	5 702	282 457	1.98	17 774	250 779
	Austria	93 537	89 073	89 073	106	88 967	0.12	4 756	82 242
	Belgium	123 469	121 493	121 209	1 324	119 885	1.09	9 690	117 912
	Canada	417 873	409 453	404 767	2 936	401 831	0.73	21 548	348 070
	Chile	274 803	252 733	252 625	2 687	249 938	1.06	6 857	229 199
	Czech Republic	96 946	93 214	93 214	1 577	91 637	1.69	6 535	82 101
	Denmark	72 310	70 854	70 854	1 965	68 889	2.77	7 481	65 642
	Estonia	12 649	12 438	12 438	442	11 996	3.55	5 867	11 634
	Finland	62 523	62 195	62 195	523	61 672	0.84	8 829	60 047
	France	792 983	755 447	755 447	27 403	728 044	3.63	5 682	701 399
	Germany	798 136	798 136	798 136	10 914	787 222	1.37	5 001	756 907
	Greece	110 521	105 096	105 096	1 364	103 732	1.30	5 125	96 640
	Hungary	111 761	108 816	108 816	1 725	107 091	1.59	4 810	91 179
	Iceland	4 505	4 491	4 491	10	4 481	0.22	3 508	4 169
	Ireland	59 296	57 979	57 952	0	57 952	0.00	5 016	54 010
	Israel	118 953	113 278	113 278	2 784	110 494	2.46	6 061	107 745
	Italy	605 490	566 973	566 973	8 498	558 475	1.50	38 142	521 288
	Japan	1 241 786	1 214 756	1 214 756	26 099	1 188 657	2.15	6 351	1 128 179
	Korea	687 104	672 101	672 101	3 053	669 048	0.45	5 033	603 632
	Luxembourg	6 187	6 082	6 082	151	5 931	2.48	5 260	5 523
	Mexico	2 114 745	1 472 875	1 472 875	7 307	1 465 568	0.50	33 806	1 326 025
	Netherlands	194 000	193 190	193 190	7 546	185 644	3.91	4 460	196 262
	New Zealand	60 940	59 118	59 118	579	58 539	0.98	5 248	53 414
	Norway	64 917	64 777	64 777	750	64 027	1.16	4 686	59 432
	Poland	425 597	410 700	410 700	6 900	403 800	1.68	5 662	379 275
Portugal	108 728	127 537	127 537	0	127 537	0.00	5 722	96 034	
Slovak Republic	59 723	59 367	59 367	1 480	57 887	2.49	5 737	54 486	
Slovenia	19 471	18 935	18 935	115	18 820	0.61	7 229	18 303	
Spain	423 444	404 374	404 374	2 031	402 343	0.50	25 335	374 266	
Sweden	102 087	102 027	102 027	1 705	100 322	1.67	4 739	94 988	
Switzerland	87 200	85 239	85 239	2 479	82 760	2.91	11 234	79 679	
Turkey	1 266 638	965 736	965 736	10 387	955 349	1.08	4 848	866 681	
United Kingdom	738 066	745 581	745 581	19 820	725 761	2.66	12 659	688 236	
United States	3 985 714	4 074 457	4 074 457	41 142	4 033 315	1.01	6 111	3 536 153	
Partners	Albania	76 910	50 157	50 157	56	50 101	0.11	4 743	42 466
	Argentina	684 879	637 603	637 603	3 995	633 608	0.63	5 908	545 942
	Brazil	3 574 928	2 786 064	2 786 064	34 932	2 751 132	1.25	20 091	2 470 804
	Bulgaria	70 188	59 684	59 684	1 437	58 247	2.41	5 282	54 255
	Colombia	889 729	620 422	620 422	4	620 418	0.00	11 173	560 805
	Costa Rica	81 489	64 326	64 326	0	64 326	0.00	4 602	40 384
	Croatia	48 155	46 550	46 550	417	46 133	0.90	6 153	45 502
	Cyprus*	9 956	9 956	9 955	128	9 827	1.29	5 078	9 650
	Hong Kong-China	84 200	77 864	77 864	813	77 051	1.04	4 670	70 636
	Indonesia	4 174 217	3 599 844	3 544 028	8 039	3 535 989	0.23	5 622	2 645 155
	Jordan	129 492	125 333	125 333	141	125 192	0.11	7 038	111 098
	Kazakhstan	258 716	247 048	247 048	7 374	239 674	2.98	5 808	208 411
	Latvia	18 789	18 389	18 375	655	17 720	3.56	5 276	16 054
	Liechtenstein	417	383	383	1	382	0.26	293	314
	Lithuania	38 524	35 567	35 567	526	35 041	1.48	4 618	33 042
	Macao-China	6 600	5 416	5 416	6	5 410	0.11	5 335	5 366
	Malaysia	544 302	457 999	457 999	225	457 774	0.05	5 197	432 080
	Montenegro	8 600	8 600	8 600	18	8 582	0.21	4 744	7 714
	Peru	584 294	508 969	508 969	263	508 706	0.05	6 035	419 945
	Qatar	11 667	11 532	11 532	202	11 330	1.75	10 966	11 003
	Romania	146 243	146 243	146 243	5 091	141 152	3.48	5 074	140 915
	Russian Federation	1 272 632	1 268 814	1 268 814	17 800	1 251 014	1.40	6 418	1 172 539
	Serbia	80 089	75 870	74 272	1 987	72 285	2.67	4 684	67 934
	Shanghai-China	108 056	90 796	90 796	1 252	89 544	1.38	6 374	85 127
	Singapore	53 637	52 163	52 163	293	51 870	0.56	5 546	51 088
	Chinese Taipei	328 356	328 336	328 336	1 747	326 589	0.53	6 046	292 542
	Thailand	982 080	784 897	784 897	9 123	775 774	1.16	6 606	703 012
	Tunisia	132 313	132 313	132 313	169	132 144	0.13	4 407	120 784
United Arab Emirates	48 824	48 446	48 446	971	47 475	2.00	11 500	40 612	
Uruguay	54 638	46 442	46 442	14	46 428	0.03	5 315	39 771	
Viet Nam	1 717 996	1 091 462	1 091 462	7 729	1 083 733	0.71	4 959	956 517	

Notes: For a full explanation of the details in this table please refer to the PISA 2012 Technical Report (OECD, forthcoming). The figure for total national population of 15-year-olds enrolled in Column 2 may occasionally be larger than the total number of 15-year-olds in Column 1 due to differing data sources. Information for the adjudicated regions is available on line.

* See notes at the beginning of this Annex.

StatLink  <http://dx.doi.org/10.1787/888933003725>



[Part 2/2]
Table A2.1 PISA target populations and samples

	Population and sample information				Coverage indices		
	Number of excluded students	Weighted number of excluded students	Within-school exclusion rate (%)	Overall exclusion rate (%)	Coverage index 1: Coverage of national desired population	Coverage index 2: Coverage of national enrolled population	Coverage index 3: Coverage of 15-year-old population
	(9)	(10)	(11)	(12)	(13)	(14)	(15)
OECD							
Australia	505	5 282	2.06	4.00	0.960	0.960	0.859
Austria	46	1 011	1.21	1.33	0.987	0.987	0.879
Belgium	39	367	0.31	1.40	0.986	0.984	0.955
Canada	1 796	21 013	5.69	6.38	0.936	0.926	0.833
Chile	18	548	0.24	1.30	0.987	0.987	0.834
Czech Republic	15	118	0.14	1.83	0.982	0.982	0.847
Denmark	368	2 381	3.50	6.18	0.938	0.938	0.908
Estonia	143	277	2.33	5.80	0.942	0.942	0.920
Finland	225	653	1.08	1.91	0.981	0.981	0.960
France	52	5 828	0.82	4.42	0.956	0.956	0.885
Germany	8	1 302	0.17	1.54	0.985	0.985	0.948
Greece	136	2 304	2.33	3.60	0.964	0.964	0.874
Hungary	27	928	1.01	2.58	0.974	0.974	0.816
Iceland	155	156	3.60	3.81	0.962	0.962	0.925
Ireland	271	2 524	4.47	4.47	0.955	0.955	0.911
Israel	114	1 884	1.72	4.13	0.959	0.959	0.906
Italy	741	9 855	1.86	3.33	0.967	0.967	0.861
Japan	0	0	0.00	2.15	0.979	0.979	0.909
Korea	17	2 238	0.37	0.82	0.992	0.992	0.879
Luxembourg	357	357	6.07	8.40	0.872	0.916	0.893
Mexico	58	3 247	0.24	0.74	0.993	0.993	0.627
Netherlands	27	1 056	0.54	4.42	0.956	0.956	1.012
New Zealand	255	2 030	3.66	4.61	0.954	0.954	0.876
Norway	278	3 133	5.01	6.11	0.939	0.939	0.916
Poland	212	11 566	2.96	4.59	0.954	0.954	0.891
Portugal	124	1 560	1.60	1.60	0.984	0.984	0.883
Slovak Republic	29	246	0.45	2.93	0.971	0.971	0.912
Slovenia	84	181	0.98	1.58	0.984	0.984	0.940
Spain	959	14 931	3.84	4.32	0.957	0.957	0.884
Sweden	201	3 789	3.84	5.44	0.946	0.946	0.930
Switzerland	256	1 093	1.35	4.22	0.958	0.958	0.914
Turkey	21	3 684	0.42	1.49	0.985	0.985	0.684
United Kingdom	486	20 173	2.85	5.43	0.946	0.946	0.932
United States	319	162 194	4.39	5.35	0.946	0.946	0.887
Partners							
Albania	1	10	0.02	0.14	0.999	0.999	0.552
Argentina	12	641	0.12	0.74	0.993	0.993	0.797
Brazil	44	4 900	0.20	1.45	0.986	0.986	0.691
Bulgaria	6	80	0.15	2.55	0.974	0.974	0.773
Colombia	23	789	0.14	0.14	0.999	0.999	0.630
Costa Rica	2	12	0.03	0.03	1.000	1.000	0.496
Croatia	91	627	1.36	2.24	0.978	0.978	0.945
Cyprus*	157	200	2.03	3.29	0.967	0.967	0.969
Hong Kong-China	38	518	0.73	1.76	0.982	0.982	0.839
Indonesia	2	860	0.03	0.26	0.997	0.982	0.634
Jordan	19	304	0.27	0.39	0.996	0.996	0.858
Kazakhstan	25	951	0.45	3.43	0.966	0.966	0.806
Latvia	14	76	0.47	4.02	0.960	0.959	0.854
Liechtenstein	13	13	3.97	4.22	0.958	0.958	0.753
Lithuania	130	867	2.56	4.00	0.960	0.960	0.858
Macao-China	3	3	0.06	0.17	0.998	0.998	0.813
Malaysia	7	554	0.13	0.18	0.998	0.998	0.794
Montenegro	4	8	0.10	0.31	0.997	0.997	0.897
Peru	8	549	0.13	0.18	0.998	0.998	0.719
Qatar	85	85	0.77	2.51	0.975	0.975	0.943
Romania	0	0	0.00	3.48	0.965	0.965	0.964
Russian Federation	69	11 940	1.01	2.40	0.976	0.976	0.921
Serbia	10	136	0.20	2.87	0.971	0.951	0.848
Shanghai-China	8	107	0.13	1.50	0.985	0.985	0.788
Singapore	33	315	0.61	1.17	0.988	0.988	0.952
Chinese Taipei	44	2 029	0.69	1.22	0.988	0.988	0.891
Thailand	12	1 144	0.16	1.32	0.987	0.987	0.716
Tunisia	5	130	0.11	0.24	0.998	0.998	0.913
United Arab Emirates	11	37	0.09	2.09	0.979	0.979	0.832
Uruguay	15	99	0.25	0.28	0.997	0.997	0.728
Viet Nam	1	198	0.02	0.73	0.993	0.993	0.557

Notes: For a full explanation of the details in this table please refer to the *PISA 2012 Technical Report* (OECD, forthcoming). The figure for total national population of 15-year-olds enrolled in Column 2 may occasionally be larger than the total number of 15-year-olds in Column 1 due to differing data sources. Information for the adjudicated regions is available on line.

* See notes at the beginning of this Annex.

StatLink <http://dx.doi.org/10.1787/888933003725>

[Part 1/1]
Table A2.2 Exclusions

	Student exclusions (unweighted)						Student exclusions (weighted)					
	Number of excluded students with functional disability (Code 1)	Number of excluded students with intellectual disability (Code 2)	Number of excluded students because of language (Code 3)	Number of excluded students for other reasons (Code 4)	Number of excluded students because of no materials available in the language of instruction (Code 5)	Total number of excluded students	Weighted number of excluded students with functional disability (Code 1)	Weighted number of excluded students with intellectual disability (Code 2)	Weighted number of excluded students because of language (Code 3)	Weighted number of excluded students for other reasons (Code 4)	Weighted number of excluded students because of no materials available in the language of instruction (Code 5)	Total weighted number of excluded students
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
OECD												
Australia	39	395	71	0	0	505	471	3 925	886	0	0	5 282
Austria	11	24	11	0	0	46	332	438	241	0	0	1 011
Belgium	5	22	12	0	0	39	24	154	189	0	0	367
Canada	82	1 593	121	0	0	1 796	981	18 682	1 350	0	0	21 013
Chile	3	15	0	0	0	18	74	474	0	0	0	548
Czech Republic	1	8	6	0	0	15	1	84	34	0	0	118
Denmark	10	204	112	42	0	368	44	1 469	559	310	0	2 381
Estonia	7	134	2	0	0	143	14	260	3	0	0	277
Finland	5	80	101	15	24	225	43	363	166	47	35	653
France	52	0	0	0	0	52	5 828	0	0	0	0	5 828
Germany	0	4	4	0	0	8	0	705	597	0	0	1 302
Greece	3	18	4	111	0	136	49	348	91	1 816	0	2 304
Hungary	1	15	2	9	0	27	36	568	27	296	0	928
Iceland	5	105	27	18	0	155	5	105	27	18	0	156
Ireland	13	159	33	66	0	271	121	1 521	283	599	0	2 524
Israel	9	91	14	0	0	114	133	1 492	260	0	0	1 884
Italy	64	566	111	0	0	741	596	7 899	1 361	0	0	9 855
Japan	0	0	0	0	0	0	0	0	0	0	0	0
Luxembourg	6	261	90	0	0	357	6	261	90	0	0	357
Mexico	21	36	1	0	0	58	812	2 390	45	0	0	3 247
Netherlands	5	21	1	0	0	27	188	819	50	0	0	1 056
New Zealand	27	118	99	0	11	255	235	926	813	0	57	2 030
Norway	11	192	75	0	0	278	120	2 180	832	0	0	3 133
Poland	23	89	6	88	6	212	1 470	5 187	177	4 644	89	11 566
Portugal	69	48	7	0	0	124	860	605	94	0	0	1 560
Korea	2	15	0	0	0	17	223	2 015	0	0	0	2 238
Slovak Republic	2	14	0	13	0	29	22	135	0	89	0	246
Slovenia	13	27	44	0	0	84	23	76	81	0	0	181
Spain	56	679	224	0	0	959	618	11 330	2 984	0	0	14 931
Sweden	120	0	81	0	0	201	2 218	0	1 571	0	0	3 789
Switzerland	7	99	150	0	0	256	41	346	706	0	0	1 093
Turkey	5	14	2	0	0	21	757	2 556	371	0	0	3 684
United Kingdom	40	405	41	0	0	486	1 468	15 514	3 191	0	0	20 173
United States	37	219	63	0	0	319	18 399	113 965	29 830	0	0	162 194
Partners												
Albania	0	0	1	0	0	1	0	0	10	0	0	10
Argentina	1	11	0	0	0	12	84	557	0	0	0	641
Brazil	17	27	0	0	0	44	1 792	3 108	0	0	0	4 900
Bulgaria	6	0	0	0	0	6	80	0	0	0	0	80
Colombia	12	10	1	0	0	23	397	378	14	0	0	789
Costa Rica	0	2	0	0	0	2	0	12	0	0	0	12
Croatia	10	78	3	0	0	91	69	539	19	0	0	627
Cyprus*	8	54	60	35	0	157	9	64	72	55	0	200
Hong Kong-China	4	33	1	0	0	38	57	446	15	0	0	518
Indonesia	1	0	1	0	0	2	426	0	434	0	0	860
Jordan	8	6	5	0	0	19	109	72	122	0	0	304
Kazakhstan	9	16	0	0	0	25	317	634	0	0	0	951
Latvia	3	7	4	0	0	14	8	45	24	0	0	76
Liechtenstein	1	7	5	0	0	13	1	7	5	0	0	13
Lithuania	10	120	0	0	0	130	66	801	0	0	0	867
Macao-China	0	1	2	0	0	3	0	1	2	0	0	3
Malaysia	3	4	0	0	0	7	274	279	0	0	0	554
Montenegro	3	1	0	0	0	4	7	1	0	0	0	8
Peru	3	5	0	0	0	8	269	280	0	0	0	549
Qatar	23	43	19	0	0	85	23	43	19	0	0	85
Romania	0	0	0	0	0	0	0	0	0	0	0	0
Russian Federation	25	40	4	0	0	69	4 345	6 934	660	0	0	11 940
Serbia	4	4	2	0	0	10	53	55	28	0	0	136
Shanghai-China	1	6	1	0	0	8	14	80	14	0	0	107
Singapore	5	17	11	0	0	33	50	157	109	0	0	315
Chinese Taipei	6	36	2	0	0	44	296	1 664	70	0	0	2 029
Thailand	2	10	0	0	0	12	13	1 131	0	0	0	1 144
Tunisia	4	1	0	0	0	5	104	26	0	0	0	130
United Arab Emirates	3	7	1	0	0	11	26	9	2	0	0	37
Uruguay	9	6	0	0	0	15	66	33	0	0	0	99
Viet Nam	0	1	0	0	0	1	0	198	0	0	0	198

Exclusion codes:

Code 1 Functional disability – student has a moderate to severe permanent physical disability.

Code 2 Intellectual disability – student has a mental or emotional disability and has either been tested as cognitively delayed or is considered in the professional opinion of qualified staff to be cognitively delayed.

Code 3 Limited assessment language proficiency – student is not a native speaker of any of the languages of the assessment in the country and has been resident in the country for less than one year.

Code 4 Other reasons defined by the national centres and approved by the international centre.

Code 5 No materials available in the language of instruction.

Note: For a full explanation of the details in this table please refer to the *PISA 2012 Technical Report* (OECD, forthcoming).

Information for the adjudicated regions is available on line.

* See notes at the beginning of this Annex.

StatLink  <http://dx.doi.org/10.1787/888933003725>



- **Column 13** presents an *index of the extent to which the national desired target population is covered by the PISA sample*. Canada, Denmark, Estonia, Luxembourg, Norway, Sweden, the United Kingdom and the United States were the only countries where the coverage is below 95%.
- **Column 14** presents an *index of the extent to which 15-year-olds enrolled in schools are covered by the PISA sample*. The index measures the overall proportion of the national enrolled population that is covered by the non-excluded portion of the student sample. The index takes into account both school-level and student-level exclusions. Values close to 100 indicate that the PISA sample represents the entire education system as defined for PISA 2012. The index is the weighted number of participating students (Column 8) divided by the weighted number of participating and excluded students (Column 8 plus Column 10), times the nationally defined target population (Column 5) divided by the eligible population (Column 2).
- **Column 15** presents an *index of the coverage of the 15-year-old population*. This index is the weighted number of participating students (Column 8) divided by the total population of 15-year-old students (Column 1).

This high level of coverage contributes to the comparability of the assessment results. For example, even assuming that the excluded students would have systematically scored worse than those who participated, and that this relationship is moderately strong, an exclusion rate in the order of 5% would likely lead to an overestimation of national mean scores of less than 5 score points (on a scale with an international mean of 500 score points and a standard deviation of 100 score points). This assessment is based on the following calculations: if the correlation between the propensity of exclusions and student performance is 0.3, resulting mean scores would likely be overestimated by 1 score point if the exclusion rate is 1%, by 3 score points if the exclusion rate is 5%, and by 6 score points if the exclusion rate is 10%. If the correlation between the propensity of exclusions and student performance is 0.5, resulting mean scores would be overestimated by 1 score point if the exclusion rate is 1%, by 5 score points if the exclusion rate is 5%, and by 10 score points if the exclusion rate is 10%. For this calculation, a model was employed that assumes a bivariate normal distribution for performance and the propensity to participate. For details, see the *PISA 2012 Technical Report* (OECD, forthcoming).

Sampling procedures and response rates

The accuracy of any survey results depends on the quality of the information on which national samples are based as well as on the sampling procedures. Quality standards, procedures, instruments and verification mechanisms were developed for PISA that ensured that national samples yielded comparable data and that the results could be compared with confidence.

Most PISA samples were designed as two-stage stratified samples (where countries applied different sampling designs, these are documented in the *PISA 2012 Technical Report* [OECD, forthcoming]). The first stage consisted of sampling individual schools in which 15-year-old students could be enrolled. Schools were sampled systematically with probabilities proportional to size, the measure of size being a function of the estimated number of eligible (15-year-old) students enrolled. A minimum of 150 schools were selected in each country (where this number existed), although the requirements for national analyses often required a somewhat larger sample. As the schools were sampled, replacement schools were simultaneously identified, in case a sampled school chose not to participate in PISA 2012.

In the case of Iceland, Liechtenstein, Luxembourg, Macao-China and Qatar, all schools and all eligible students within schools were included in the sample.

Experts from the PISA Consortium performed the sample selection process for most participating countries and monitored it closely in those countries that selected their own samples. The second stage of the selection process sampled students within sampled schools. Once schools were selected, a list of each sampled school's 15-year-old students was prepared. From this list, 35 students were then selected with equal probability (all 15-year-old students were selected if fewer than 35 were enrolled). The number of students to be sampled per school could deviate from 35, but could not be less than 20.

Data-quality standards in PISA required minimum participation rates for schools as well as for students. These standards were established to minimise the potential for response biases. In the case of countries meeting these standards, it was likely that any bias resulting from non-response would be negligible, i.e. typically smaller than the sampling error.

A minimum response rate of 85% was required for the schools initially selected. Where the initial response rate of schools was between 65% and 85%, however, an acceptable school response rate could still be achieved through the use of replacement schools. This procedure brought with it a risk of increased response bias. Participating countries were, therefore, encouraged to persuade as many of the schools in the original sample as possible to participate. Schools with a student participation rate between 25% and 50% were not regarded as participating schools, but data from these schools were included in the database and contributed to the various estimations. Data from schools with a student participation rate of less than 25% were excluded from the database.

PISA 2012 also required a minimum participation rate of 80% of students within participating schools. This minimum participation rate had to be met at the national level, not necessarily by each participating school. Follow-up sessions were required in schools in which too few students had participated in the original assessment sessions. Student participation rates were calculated over all original schools, and also over all schools, whether original sample or replacement schools, and from the participation of students in both the original assessment and any follow-up sessions. A student who participated in the original or follow-up cognitive sessions was regarded as a participant. Those who attended only the questionnaire session were included in the international database and contributed to the statistics presented in this publication if they provided at least a description of their father's or mother's occupation.

[Part 1/2]

Table A2.3 Response rates

	Initial sample – before school replacement					Final sample – after school replacement		
	Weighted school participation rate before replacement (%)	Weighted number of responding schools (weighted also by enrolment)	Weighted number of schools sampled (responding and non-responding) (weighted also by enrolment)	Number of responding schools (unweighted)	Number of non-responding and non-responding schools (unweighted)	Weighted school participation rate after replacement (%)	Weighted number of responding schools (weighted also by enrolment)	Weighted number of schools sampled (responding and non-responding) (weighted also by enrolment)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
OECD								
Australia	98	268 631	274 432	757	790	98	268 631	274 432
Austria	100	88 967	88 967	191	191	100	88 967	88 967
Belgium	84	100 482	119 019	246	294	97	115 004	119 006
Canada	91	362 178	396 757	828	907	93	368 600	396 757
Chile	92	220 009	239 429	200	224	99	236 576	239 370
Czech Republic	98	87 238	88 884	292	297	100	88 447	88 797
Denmark	87	61 749	71 015	311	366	96	67 709	70 892
Estonia	100	12 046	12 046	206	206	100	12 046	12 046
Finland	99	59 740	60 323	310	313	99	59 912	60 323
France	97	703 458	728 401	223	231	97	703 458	728 401
Germany	98	735 944	753 179	227	233	98	737 778	753 179
Greece	93	95 107	102 087	176	192	99	100 892	102 053
Hungary	98	99 317	101 751	198	208	99	101 187	101 751
Iceland	99	4 395	4 424	133	140	99	4 395	4 424
Ireland	99	56 962	57 711	182	185	99	57 316	57 711
Israel	91	99 543	109 326	166	186	94	103 075	109 895
Italy	89	478 317	536 921	1 104	1 232	97	522 686	536 821
Japan	86	1 015 198	1 175 794	173	200	96	1 123 211	1 175 794
Korea	100	661 575	662 510	156	157	100	661 575	662 510
Luxembourg	100	5 931	5 931	42	42	100	5 931	5 931
Mexico	92	1 323 816	1 442 242	1 431	1 562	95	1 374 615	1 442 234
Netherlands	75	139 709	185 468	148	199	89	165 635	185 320
New Zealand	81	47 441	58 676	156	197	89	52 360	58 616
Norway	85	54 201	63 653	177	208	95	60 270	63 642
Poland	85	343 344	402 116	159	188	98	393 872	402 116
Portugal	95	122 238	128 129	186	195	96	122 713	128 050
Slovak Republic	87	50 182	57 353	202	236	99	57 599	58 201
Slovenia	98	18 329	18 680	335	353	98	18 329	18 680
Spain	100	402 604	403 999	902	904	100	402 604	403 999
Sweden	99	98 645	99 726	207	211	100	99 536	99 767
Switzerland	94	78 825	83 450	397	422	98	82 032	83 424
Turkey	97	921 643	945 357	165	170	100	944 807	945 357
United Kingdom	80	564 438	705 011	477	550	89	624 499	699 839
United States	67	2 647 253	3 945 575	139	207	77	3 040 661	3 938 077
Partners								
Albania	100	49 632	49 632	204	204	100	49 632	49 632
Argentina	95	578 723	606 069	218	229	96	580 989	606 069
Brazil	93	2 545 863	2 745 045	803	886	95	2 622 293	2 747 688
Bulgaria	99	57 101	57 574	186	188	100	57 464	57 574
Colombia	87	530 553	612 605	323	363	97	596 557	612 261
Costa Rica	99	64 235	64 920	191	193	99	64 235	64 920
Croatia	99	45 037	45 636	161	164	100	45 608	45 636
Cyprus*	97	9 485	9 821	117	131	97	9 485	9 821
Hong Kong-China	79	60 277	76 589	123	156	94	72 064	76 567
Indonesia	95	2 799 943	2 950 696	199	210	98	2 892 365	2 951 028
Jordan	100	119 147	119 147	233	233	100	119 147	119 147
Kazakhstan	100	239 767	239 767	218	218	100	239 767	239 767
Latvia	88	15 371	17 488	186	213	100	17 428	17 448
Liechtenstein	100	382	382	12	12	100	382	382
Lithuania	98	33 989	34 614	211	216	100	34 604	34 604
Macao-China	100	5 410	5 410	45	45	100	5 410	5 410
Malaysia	100	455 543	455 543	164	164	100	455 543	455 543
Montenegro	100	8 540	8 540	51	51	100	8 540	8 540
Peru	98	503 915	514 574	238	243	99	507 602	514 574
Qatar	100	11 333	11 340	157	164	100	11 333	11 340
Romania	100	139 597	139 597	178	178	100	139 597	139 597
Russian Federation	100	1 243 564	1 243 564	227	227	100	1 243 564	1 243 564
Serbia	90	65 537	72 819	143	160	95	69 433	72 752
Shanghai-China	100	89 832	89 832	155	155	100	89 832	89 832
Singapore	98	50 415	51 687	170	176	98	50 945	51 896
Chinese Taipei	100	324 667	324 667	163	163	100	324 667	324 667
Thailand	98	757 516	772 654	235	240	100	772 452	772 654
Tunisia	99	129 229	130 141	152	153	99	129 229	130 141
United Arab Emirates	99	46 469	46 748	453	460	99	46 469	46 748
Uruguay	99	45 736	46 009	179	180	100	46 009	46 009
Viet Nam	100	1 068 462	1 068 462	162	162	100	1 068 462	1 068 462

Information for the adjudicated regions is available on line.

* See notes at the beginning of this Annex.

StatLink  <http://dx.doi.org/10.1787/888933003725>



[Part 2/2]
Table A2.3 Response rates

	Final sample – after school replacement		Final sample – students within schools after school replacement				
	Number of responding schools (unweighted)	Number of non-responding schools (unweighted)	Weighted student participation rate after replacement (%)	Number of students assessed (weighted)	Number of students sampled (assessed and absent) (weighted)	Number of students assessed (unweighted)	Number of students sampled (assessed and absent) (unweighted)
	(9)	(10)	(11)	(12)	(13)	(14)	(15)
OECD							
Australia	757	790	87	213 495	246 012	17 491	20 799
Austria	191	191	92	75 393	82 242	4 756	5 318
Belgium	282	294	91	103 914	114 360	9 649	10 595
Canada	840	907	81	261 928	324 328	20 994	25 835
Chile	221	224	95	214 558	226 689	6 857	7 246
Czech Republic	295	297	90	73 536	81 642	6 528	7 222
Denmark	339	366	89	56 096	62 988	7 463	8 496
Estonia	206	206	93	10 807	11 634	5 867	6 316
Finland	311	313	91	54 126	59 653	8 829	9 789
France	223	231	89	605 371	676 730	5 641	6 308
Germany	228	233	93	692 226	742 416	4 990	5 355
Greece	188	192	97	92 444	95 580	5 125	5 301
Hungary	204	208	93	84 032	90 652	4 810	5 184
Iceland	133	140	85	3 503	4 135	3 503	4 135
Ireland	183	185	84	45 115	53 644	5 016	5 977
Israel	172	186	90	91 181	101 288	6 061	6 727
Italy	1 186	1 232	93	473 104	510 005	38 084	41 003
Japan	191	200	96	1 034 803	1 076 786	6 351	6 609
Korea	156	157	99	595 461	603 004	5 033	5 101
Luxembourg	42	42	95	5 260	5 523	5 260	5 523
Mexico	1 468	1 562	94	1 193 866	1 271 639	33 786	35 972
Netherlands	177	199	85	148 432	174 697	4 434	5 215
New Zealand	177	197	85	40 397	47 703	5 248	6 206
Norway	197	208	91	51 155	56 286	4 686	5 156
Poland	182	188	88	325 389	371 434	5 629	6 452
Portugal	187	195	87	80 719	92 395	5 608	6 426
Slovak Republic	231	236	94	50 544	53 912	5 737	6 106
Slovenia	335	353	90	16 146	17 849	7 211	7 921
Spain	902	904	90	334 382	372 042	26 443	29 027
Sweden	209	211	92	87 359	94 784	4 739	5 141
Switzerland	410	422	92	72 116	78 424	11 218	12 138
Turkey	169	170	98	850 830	866 269	4 847	4 939
United Kingdom	505	550	86	528 231	613 736	12 638	14 649
United States	161	207	89	2 429 718	2 734 268	6 094	6 848
Partners							
Albania	204	204	92	39 275	42 466	4 743	5 102
Argentina	219	229	88	457 294	519 733	5 804	6 680
Brazil	837	886	90	2 133 035	2 368 438	19 877	22 326
Bulgaria	187	188	96	51 819	54 145	5 280	5 508
Colombia	352	363	93	507 178	544 862	11 164	12 045
Costa Rica	191	193	89	35 525	39 930	4 582	5 187
Croatia	163	164	92	41 912	45 473	6 153	6 675
Cyprus*	117	131	93	8 719	9 344	5 078	5 458
Hong Kong-China	147	156	93	62 059	66 665	4 659	5 004
Indonesia	206	210	95	2 478 961	2 605 254	5 579	5 885
Jordan	233	233	95	105 493	111 098	7 038	7 402
Kazakhstan	218	218	99	206 053	208 411	5 808	5 874
Latvia	211	213	91	14 579	16 039	5 276	5 785
Liechtenstein	12	12	93	293	314	293	314
Lithuania	216	216	92	30 429	33 042	4 618	5 018
Macao-China	45	45	99	5 335	5 366	5 335	5 366
Malaysia	164	164	94	405 983	432 080	5 197	5 529
Montenegro	51	51	94	7 233	7 714	4 799	5 117
Peru	240	243	96	398 193	414 728	6 035	6 291
Qatar	157	164	100	10 966	10 996	10 966	10 996
Romania	178	178	98	137 860	140 915	5 074	5 188
Russian Federation	227	227	97	1 141 317	1 172 539	6 418	6 602
Serbia	152	160	93	60 366	64 658	4 681	5 017
Shanghai-China	155	155	98	83 821	85 127	6 374	6 467
Singapore	172	176	94	47 465	50 330	5 546	5 887
Chinese Taipei	163	163	96	281 799	292 542	6 046	6 279
Thailand	239	240	99	695 088	702 818	6 606	6 681
Tunisia	152	153	90	108 342	119 917	4 391	4 857
United Arab Emirates	453	460	95	38 228	40 384	11 460	12 148
Uruguay	180	180	90	35 800	39 771	5 315	5 904
Viet Nam	162	162	100	955 222	956 517	4 959	4 966

Information for the adjudicated regions is available on line.

* See notes at the beginning of this Annex.

StatLink  <http://dx.doi.org/10.1787/888933003725>



Table A2.3 shows the response rates for students and schools, before and after replacement.

- **Column 1** shows the *weighted participation rate of schools before replacement*. This is obtained by dividing Column 2 by Column 3, multiply by 100.
- **Column 2** shows the *weighted number of responding schools before school replacement* (weighted by student enrolment).
- **Column 3** shows the *weighted number of sampled schools before school replacement* (including both responding and non-responding schools, weighted by student enrolment).
- **Column 4** shows the *unweighted number of responding schools before school replacement*.
- **Column 5** shows the *unweighted number of responding and non-responding schools before school replacement*.
- **Column 6** shows the *weighted participation rate of schools after replacement*. This is obtained by dividing Column 7 by Column 8, multiply by 100.
- **Column 7** shows the *weighted number of responding schools after school replacement* (weighted by student enrolment).
- **Column 8** shows the *weighted number of schools sampled after school replacement* (including both responding and non-responding schools, weighted by student enrolment).
- **Column 9** shows the *unweighted number of responding schools after school replacement*.
- **Column 10** shows the *unweighted number of responding and non-responding schools after school replacement*.
- **Column 11** shows the *weighted student participation rate after replacement*. This is obtained by dividing Column 12 by Column 13, multiply by 100.
- **Column 12** shows the *weighted number of students assessed*.
- **Column 13** shows the *weighted number of students sampled* (including both students who were assessed and students who were absent on the day of the assessment).
- **Column 14** shows the *unweighted number of students assessed*. Note that any students in schools with student-response rates less than 50% were not included in these rates (both weighted and unweighted).
- **Column 15** shows the *unweighted number of students sampled* (including both students that were assessed and students who were absent on the day of the assessment). Note that any students in schools where fewer than half of the eligible students were assessed were not included in these rates (neither weighted nor unweighted).

Differences between the problem-solving sample and the main PISA student sample

Out of the 65 countries and economies that participated in PISA 2012, 44 also implemented the computer-based assessment (CBA) of problem solving. Of these, 12 countries and economies only assessed problem solving, while 32 also assessed mathematics and (digital) reading on computers.

In all 44 countries/economies, only a random sub-sample of students who participated in the paper-based assessment (PBA) of mathematics were sampled to be administered the assessment of problem solving. However, as long as at least one student in a participating school was sampled for the computer-based assessment, all students in the PISA sample from that school received multiple imputations (plausible values) of performance in problem solving. This is similar to the procedure used to impute plausible values for minor domains in PISA (for instance, not all test booklets in 2012 included reading questions; but all students received imputed values for reading performance).

Table A2.4 compares the final samples (after school replacement) for mathematics and problem solving.

- **Column 1** shows the overall number of schools with valid data in the PISA 2012 database.
- **Column 2** shows the students with valid data in mathematics. This is the number of students with data included in the main database. All these students have imputed values for performance in mathematics, reading and science. Students are considered as participating in the assessment of mathematics if they were sampled to sit the paper-based assessment (all booklets included mathematics questions) and attended a test session. Those who only attended the questionnaire session but provided at least a description of their father's or mother's occupation are also regarded as participants.
- **Column 3** shows the number of schools with valid data in the PISA 2012 computer-based assessments database.
- **Column 4** shows the number of students with valid data in problem solving. This corresponds to all participating students (Column 2) within schools who were sampled for the computer-based assessments in PISA 2012 and were included in the database (Column 3). For all these students, performance in problem solving could be imputed. All these students contributed to the statistics presented in this publication (with the exception of statistics based on item-level performance).
- **Column 5** shows the number of students included in the database who were sampled for the assessment of problem solving. These are the students with valid data who were sampled to sit the computer-based assessment and assigned a form (the computer equivalent of a paper booklet) containing at least one cluster of problem-solving questions.
- **Column 6** shows the number of students who were actually assessed in problem solving. These are the students sampled for the assessment of problem solving who actually attended the computer-based assessment session and were administered the test. All these students contributed to statistics based on item-level performance in this volume. Differences between the number of students in Columns 5 and 6 can occur for several reasons: students who skipped the computer-based session; students who did not reach any of the problem-solving questions in their test form; technical problems with the computer; etc.



[Part 1/1]
Table A2.4 Sample size for performance in mathematics and problem solving

	Mathematics		Problem solving			
	Number of schools with valid data (unweighted)	Number of students with valid data (unweighted)	Number of schools with valid data (unweighted)	Number of students with valid data (unweighted)	Number of students with valid data sampled for the assessment of problem solving (unweighted)	Number of students who were administered the assessment of problem solving (unweighted)
	(1)	(2)	(3)	(4)	(5)	(6)
OECD						
Australia	775	14 481	775	14 481	5 922	5 612
Austria	191	4 755	191	4 755	1 376	1 331
Belgium	287	8 597	287	8 597	2 309	2 147
Canada	885	21 544	885	21 544	5 415	4 602
Chile	221	6 856	221	6 856	1 674	1 578
Czech Republic	297	5 327	297	5 327	3 229	3 076
Denmark	341	7 481	341	7 481	2 104	1 948
Estonia	206	4 779	206	4 779	1 412	1 367
Finland	311	8 829	311	8 829	3 685	3 531
France	226	4 613	226	4 613	1 509	1 345
Germany	230	5 001	230	5 001	1 426	1 350
Greece	188	5 125	0	0	0	0
Hungary	204	4 810	204	4 810	1 355	1 300
Iceland	134	3 508	0	0	0	0
Ireland	183	5 016	183	5 016	1 303	1 190
Israel	172	5 055	172	5 055	1 445	1 346
Italy	1 194	31 073	208	5 495	1 554	1 371
Japan	191	6 351	191	6 351	3 178	3 014
Korea	156	5 033	156	5 033	1 351	1 336
Luxembourg	42	5 258	0	0	0	0
Mexico	1 471	33 806	0	0	0	0
Netherlands	179	4 460	179	4 460	2 258	1 752
New Zealand	177	4 291	0	0	0	0
Norway	197	4 686	197	4 686	1 463	1 240
Poland	184	4 607	184	4 607	1 256	1 227
Portugal	195	5 722	195	5 722	1 631	1 446
Slovak Republic	231	4 678	231	4 678	1 589	1 465
Slovenia	338	5 911	338	5 911	2 179	2 065
Spain	902	25 313	368	10 175	2 866	2 709
Sweden	209	4 736	209	4 736	1 337	1 258
Switzerland	411	11 229	0	0	0	0
Turkey	170	4 848	170	4 848	2 022	1 995
United Kingdom	507	12 659	170	4 185	1 963	1 458
United States	162	4 978	162	4 978	1 300	1 273
Partners						
Albania	204	4 743	0	0	0	0
Argentina	226	5 908	0	0	0	0
Brazil	839	19 204	241	5 506	1 590	1 463
Bulgaria	188	5 282	188	5 282	2 333	2 145
Colombia	352	9 073	352	9 073	2 595	2 307
Costa Rica	193	4 602	0	0	0	0
Croatia	163	5 008	163	5 008	2 016	1 924
Cyprus*	117	5 078	117	5 078	2 630	2 503
Hong Kong-China	148	4 670	148	4 670	1 367	1 325
Indonesia	209	5 622	0	0	0	0
Jordan	233	7 038	0	0	0	0
Kazakhstan	218	5 808	0	0	0	0
Latvia	211	4 306	0	0	0	0
Liechtenstein	12	293	0	0	0	0
Lithuania	216	4 618	0	0	0	0
Macao-China	45	5 335	45	5 335	1 577	1 565
Malaysia	164	5 197	164	5 197	2 072	1 929
Montenegro	51	4 744	51	4 744	2 101	1 845
Peru	240	6 035	0	0	0	0
Qatar	157	10 966	0	0	0	0
Romania	178	5 074	0	0	0	0
Russian Federation	227	5 231	227	5 231	1 574	1 543
Serbia	153	4 684	153	4 684	1 930	1 777
Shanghai-China	155	5 177	155	5 177	1 213	1 203
Singapore	172	5 546	172	5 546	1 438	1 394
Chinese Taipei	163	6 046	163	6 046	1 512	1 484
Thailand	239	6 606	0	0	0	0
Tunisia	153	4 407	0	0	0	0
United Arab Emirates	458	11 500	458	11 500	3 418	3 262
Uruguay	180	5 315	180	5 315	2 048	2 013
Viet Nam	162	4 959	0	0	0	0

* See notes at the beginning of this Annex.

StatLink  <http://dx.doi.org/10.1787/888933003725>



In all but four of the 44 countries/economies that assessed problem solving, the school samples for CBA and PBA coincide. As a consequence, in 40 countries/economies the main student dataset, containing the results of paper-based assessments, and the CBA dataset have the same number of observations. In Brazil, Italy, Spain and the United Kingdom, in contrast, the CBA school sample is smaller than the main sample. Brazil and Italy did not over-sample students for CBA to provide results at regional level. In Spain, students were over-sampled only in the Basque Country and in Catalonia, but not in the remaining adjudicated regions. In the United Kingdom, only schools in England participated in the computer-based assessment of problem solving.

Definition of schools

In some countries, sub-units within schools were sampled instead of schools and this may affect the estimation of the between-school variance components. In Austria, the Czech Republic, Germany, Hungary, Japan, Romania and Slovenia, schools with more than one study programme were split into the units delivering these programmes. In the Netherlands, for schools with both lower and upper secondary programmes, schools were split into units delivering each programme level. In the Flemish Community of Belgium, in the case of multi-campus schools, implantations (campuses) were sampled, whereas in the French Community, in the case of multi-campus schools, the larger administrative units were sampled. In Australia, for schools with more than one campus, the individual campuses were listed for sampling. In Argentina, Croatia and Dubai (United Arab Emirates), schools that had more than one campus had the locations listed for sampling. In Spain, the schools in the Basque region with multi-linguistic models were split into linguistic models for sampling.



ANNEX A3

TECHNICAL NOTES ON ANALYSES IN THIS VOLUME

Methods and definitions

Relative performance in problem solving

Relative performance in problem solving is defined as the difference between a student's actual performance in problem solving and his or her expected performance, based on performance in other domains:

$$RP_i^{ps} = y_i^{ps} - E(y_i^{ps} | y_i^{mrs})$$

where y_i^{ps} represents student i 's performance in problem solving, and y_i^{mrs} is a vector of student i 's performance in other domains (such as mathematics, reading and science).

A student's (conditionally) expected performance is estimated using regression models; relative performance is therefore based on residuals from regression models. All analyses of relative performance in this volume derive residuals from parametric regression models that allow for curvilinear shapes and, when more than one domain enters the conditioning arguments, for interaction terms (second- or third-degree polynomials). However, different regression methods can be used, including non-parametric ones. Figure V.2.16, for instance, graphically displays a non-parametric regression of problem-solving performance on mathematics performance.

In some analyses, the regression model is calibrated only on a subsample of comparison students (e.g. on boys, when the relative performance of girls is analysed). In others, where the comparison group is less well defined and the focus is on comparisons to the national or international average, the regression model is calibrated on all students. In all cases, five distinct regression models are estimated to compute five plausible values of relative performance.

Relative risk or increased likelihood

The relative risk is a measure of the association between an antecedent factor and an outcome factor. The relative risk is simply the ratio of two risks, i.e. the risk of observing the outcome when the antecedent is present and the risk of observing the outcome when the antecedent is not present. Figure A3.1 presents the notation that is used in the following.

■ Figure A3.1 ■

Labels used in a two-way table

P_{11}	P_{12}	$P_{1.}$
P_{21}	P_{22}	$P_{2.}$
$P_{.1}$	$P_{.2}$	$P_{..}$

$P_{.}$ is equal to $\frac{n_{.}}{n_{..}}$, with $n_{.}$ the total number of students and $P_{.}$ is therefore equal to 1, $P_{i.}$, $P_{.j}$ respectively represent the marginal probabilities for each row and for each column. The marginal probabilities are equal to the marginal frequencies divided by the total number of students. Finally, the P_{ij} represents the probabilities for each cell and are equal to the number of observations in a particular cell divided by the total number of observations.

In PISA, the rows represent the antecedent factor, with the first row for "having the antecedent" and the second row for "not having the antecedent". The columns represent the outcome: the first column for "having the outcome" and the second column for "not having the outcome". The relative risk is then equal to:

$$RR = \frac{(P_{11}/P_{1.})}{(P_{21}/P_{2.})}$$

Statistics based on multilevel models

Statistics based on multilevel models include variance components (between- and within-school variance), the *index of inclusion* derived from these components, and regression coefficients where this has been indicated. Multilevel models are generally specified as two-level regression models (the student and school levels), with normally distributed residuals, and estimated with maximum likelihood estimation. Where the dependent variable is mathematics performance, the estimation uses five plausible values for each student's performance on the mathematics scale. Models were estimated using Mplus® software.

In multilevel models, weights are used at both the student and school levels. The purpose of these weights is to account for differences in the probabilities of students being selected in the sample. Since PISA applies a two-stage sampling procedure, these differences are due to factors at both the school and the student levels. For the multilevel models, student final weights (W_FSTUWT) were used.



Within-school-weights correspond to student final weights, rescaled to sum up within each school to the school sample size. Between-school weights correspond to the sum of student final weights (W_FSTUWT) within each school. The definition of between-school weights has changed with respect to PISA 2009.

The *index of inclusion* is defined and estimated as:

$$100 * \frac{\sigma_w^2}{\sigma_w^2 + \sigma_b^2}$$

where σ_w^2 and σ_b^2 , respectively, represent the within- and between-variance estimates.

The results in multilevel models, and the between-school variance estimate in particular, depend on how schools are defined and organised within countries and by the units that were chosen for sampling purposes. For example, in some countries, some of the schools in the PISA sample were defined as administrative units (even if they spanned several geographically separate institutions, as in Italy); in others they were defined as those parts of larger educational institutions that serve 15-year-olds; in still others they were defined as physical school buildings; and in others they were defined from a management perspective (e.g. entities having a principal). The *PISA 2012 Technical Report* (OECD, forthcoming) and Annex A2 provide an overview of how schools were defined. In Slovenia, the primary sampling unit is defined as a group of students who follow the same study programme within a school (an educational track within a school). So in this particular case the between-school variance is actually the within-school, between-track variation. The use of stratification variables in the selection of schools may also affect the estimate of the between-school variance, particularly if stratification variables are associated with between-school differences.

Because of the manner in which students were sampled, the within-school variation includes variation between classes as well as between students.

Effect sizes

An effect size is a measure of the strength of the relationship between two variables. The term effect size is commonly used to refer to standardised differences. Standardising a difference is useful when a metric has no intrinsic meaning – as is the case with PISA performance scales or scale indices. Indeed, a standardised difference allows comparisons of the strength of between-group differences across measures that vary in their metric.

A standardised difference is obtained by dividing the raw difference between two groups, such as boys and girls, by a measure of the variation in the underlying data. In this volume, the pooled standard deviation was used to standardise differences. The effect size between two subgroups is thus calculated as:

$$\frac{m_1 - m_2}{\sqrt{\sigma_{1,2}^2}}$$

where m_1 and m_2 , respectively, represent the mean values for the subgroups 1 and 2, and $\sigma_{1,2}^2$ represents the variance for the population pooling subgroups 1 and 2.

Relative success ratios on subsets of items

The relative likelihood of success on a subset of items is computed as follows.

First, a country-specific measure of success on each item is computed by converting the percentage of correct answers into the logit scale (the logarithm of odds is used instead of the percentage; odds are also referred to as success ratios, because they correspond to the number of full-credit answers over the number of no- and partial-credit answers). This success measure can also be interpreted as an item-difficulty parameter: lower success measures indicate more difficult items.

Next, a relative success measure for a given subset of items is derived as the difference between the average success on items in the subset and the average success on items outside of the subset. Again, this measure can also be interpreted as a relative difficulty of items in the two subsets.

Finally, a relative likelihood of success is derived that takes into account differences in item difficulty by subtracting the average relative success in OECD countries (i.e. the average difficulty of items) from country-specific figures (or similarly, the relative success in a comparison group – e.g. boys – from the relative success in the focus group – e.g. girls). This difference is used as a basis for computing odds ratios (the difference of logits being the logarithm of the odds ratio).

By design, each item carries the same weight in these analyses. However, the probability of success on a given item is also influenced by its position within the test booklet. While *ex ante*, booklets are assigned so that they are present in equal proportions within any subsample, in practice given the finite number of students taking the test small differences remain. To control for these differences, booklet dummies are included in the model and generalised odds ratios are estimated with logistic regression. Similarly, in some analyses country- or group-specific dummies are included for the response format to ensure that inferences about strengths and weaknesses on the items measuring the various framework aspects are not driven by the association of selected- and constructed-response formats with specific item families.



Standard errors and significance tests

The statistics in this report represent estimates of national performance based on samples of students, rather than values that could be calculated if every student in every country had answered every question. Consequently, it is important to measure the degree of uncertainty of the estimates. In PISA, each estimate has an associated degree of uncertainty, which is expressed through a standard error. The use of confidence intervals provides a way to make inferences about the population means and proportions in a manner that reflects the uncertainty associated with the sample estimates. From an observed sample statistic and assuming a normal distribution, it can be inferred that the corresponding population result would lie within the confidence interval in 95 out of 100 replications of the measurement on different samples drawn from the same population.

In many cases, readers are primarily interested in whether a given value in a particular country is different from a second value in the same or another country, e.g. whether girls in a country perform better than boys in the same country. In the tables and charts used in this report, differences are labelled as statistically significant when a difference of that size, smaller or larger, would be observed less than 5% of the time, if there were actually no difference in corresponding population values. Similarly, the risk of reporting a correlation as significant if there is, in fact, no correlation between two measures, is contained at 5%.

Throughout the report, significance tests were undertaken to assess the statistical significance of the comparisons made.

Gender differences and differences between subgroup means

Gender differences in student performance or other indices were tested for statistical significance. Positive differences indicate higher scores for boys while negative differences indicate higher scores for girls. Generally, differences marked in bold in the tables in this volume are statistically significant at the 95% confidence level.

Similarly, differences between other groups of students (e.g. native students and students with an immigrant background) were tested for statistical significance. The definitions of the subgroups can in general be found in the tables and the text accompanying the analysis. All differences marked in bold in the tables presented in Annex B of this report are statistically significant at the 95% level.

Differences between subgroup means, after accounting for other variables

For many tables, subgroup comparisons were performed both on the observed difference (“before accounting for other variables”) and after accounting for other variables, such as the *PISA index of economic, social and cultural status of students* (ESCS). The adjusted differences were estimated using linear regression and tested for significance at the 95% confidence level. Significant differences are marked in bold.

Performance differences between the top and bottom quartiles of PISA indices and scales

Differences in average performance between the top and bottom quarters of the PISA indices and scales were tested for statistical significance. Figures marked in bold indicate that performance between the top and bottom quarters of students on the respective index is statistically significantly different at the 95% confidence level.

Change in the performance per unit of the index

For many tables, the difference in student performance per unit of the index shown was calculated. Figures in bold indicate that the differences are statistically significantly different from zero at the 95% confidence level.

Relative risk or increased likelihood

Figures in bold in the data tables presented in Annex B of this report indicate that the relative risk is statistically significantly different from 1 at the 95% confidence level. To compute statistical significance around the value of 1 (the null hypothesis), the relative-risk statistic is assumed to follow a log-normal distribution, rather than a normal distribution, under the null hypothesis.

Range of ranks

To calculate the range of ranks for countries, data are simulated using the mean and standard error of the mean for each relevant country to generate a distribution of possible values. Some 10 000 simulations are implemented and, based on these values, 10 000 possible rankings for each country are produced. For each country, the counts for each rank are aggregated from largest to smallest until they equal 9 500 or more. Then the range of ranks per country is reported, including all the ranks that have been aggregated. This means that there is at least 95% confidence about the range of ranks, and it is safe to assume unimodality in this distribution of ranks. This method has been used in all cycles of PISA since 2003, including PISA 2012.

The main difference between the range of ranks (e.g. Figure V.2.4) and the comparison of countries’ mean performance (e.g. Figure V.2.3) is that the former takes account of the multiple comparisons involved in determining ranks and the asymmetry of the distribution of rank estimates, while the latter does not. Therefore, sometimes there is a slight difference between the range of ranks and counting the number of countries above a given country, based on pairwise comparisons of the selected countries’ performance. For instance, the difference in average performance between England (United Kingdom), which is listed in eleventh place in Figure V.2.3, and Canada, which is listed in eighth place, is not statistically significant. However, because it is highly unlikely that all three countries/economies listed between eighth and tenth place in reality have lower performance than England (United Kingdom), the rank for England (United Kingdom)



among all countries can be restricted to be, with 95% confidence, at best ninth (Figure V.2.4). Since it is safe to assume that the distribution of rank estimates for each country has a single mode (unimodality), the results of range of ranks for countries should be used when examining countries' rankings.

Standard errors in statistics estimated from multilevel models

For statistics based on multilevel models (such as the estimates of variance components and regression coefficients from two-level regression models) the standard errors are not estimated with the usual replication method which accounts for stratification and sampling rates from finite populations. Instead, standard errors are "model-based": their computation assumes that schools, and students within schools, are sampled at random (with sampling probabilities reflected in school and student weights) from a theoretical, infinite population of schools and students which complies with the model's parametric assumptions.

The standard error for the estimated *index of inclusion* is calculated by deriving an approximate distribution for it from the (model-based) standard errors for the variance components, using the delta-method.

Differences between rankings based on proficiency scales and average percent-correct rankings

PISA international results are based on a scaling of students' item scores with an item response model (see the *PISA 2012 Technical Report*, OECD, forthcoming). This scaling is undertaken for a number of reasons. First, it supports the construction of described proficiency scales. Second, this approach summarises students' responses to many items with few indices. In doing so, it ensures that the indices are comparable across students who respond to different test booklets that are composed of different subsets of items (Adams et al., 2010). The scaling of students' scores reflects the PISA approach, which consists in building internationally supported assessment frameworks and then developing items pools that sample widely from those frameworks in an agreed fashion.

The average percent-correct approach used in Chapter 3 in this volume provides an alternative way of comparing country performance on the assessment. The advantage of the average percent-correct approach is that it can be easily replicated on arbitrary subsets of items.

When rankings based on the percent-correct approach, using all items, are compared to rankings based on the usual scaling approach, small differences will occur for six reasons. First, the percent-correct methodology assigns an arbitrary value (typically, either 0 or 0.5) to all partial-credit answers; percent-correct figures are therefore based on a smaller set of information about students' performance on the test than scaled results, where each partial credit value is scaled to its specific difficulty. Second, the percent-correct methodology ignores students who did not answer any problem-solving item, despite being assigned to a problem-solving booklet and having answered, at least partially, the student questionnaire. Because it is impossible to know why they did not answer problem-solving questions (e.g. a technical failure of the computer system or a deliberate absence from the test), their answers are coded as "not administered" rather than as incorrect, and treated as missing. The usual scaling approach, in contrast, corrects for possible self-selection in taking the test by imputing performance from the available information about these students, including their performance on other tests. Third, the percent-correct methodology weights all items equally, whereas in the scaling approach the items are weighted according to the number of booklets in which they were included. Fourth, the percent-correct approach does not address the booklet effect that was observed in PISA. Fifth, the scaling methodology transforms percentage values that are bounded at zero and 100 into the logit scale. This transformation has the effect of "stretching out" very low and very high percentages in comparison to percentages that are close to 50%. Sixth, when a problem such as a translation error affecting one item in one country is detected after the test has been administered, this item is coded as missing for all students in the country; the percent-correct rankings may therefore be based on fewer items than the scaled results. In the PISA 2012 assessment of problem solving, one item (CP018Q05) was withdrawn after the test in France, because by mistake a crucial direction to students had not been included in the national version.

References

- Adams, R., A. Berezner and M. Jakubowski (2010), "Analysis of PISA 2006 Preferred Items Ranking Using the Percent-Correct Method", *OECD Education Working Papers*, No. 46, OECD Publishing.
<http://dx.doi.org/10.1787/5km4psmntkq5-e>
- OECD (forthcoming), *PISA 2012 Technical Report*, PISA, OECD Publishing.



ANNEX A4

QUALITY ASSURANCE

Quality assurance procedures were implemented in all parts of PISA 2012, as was done for all previous PISA surveys.

The consistent quality and linguistic equivalence of the PISA 2012 assessment instruments were facilitated by providing countries with equivalent source versions of the assessment instruments in English and French and requiring countries (other than those assessing students in English and French) to prepare and consolidate two independent translations using both source versions. Precise translation and adaptation guidelines were supplied, also including instructions for selecting and training the translators. For each country, the translation and format of the assessment instruments (including test materials, marking guides, questionnaires and manuals) were verified by expert translators appointed by the PISA Consortium before they were used in the PISA 2012 field trial and main study. These translators' mother tongue was the language of instruction in the country concerned and they were knowledgeable about education systems. For further information on the PISA translation procedures, see the *PISA 2012 Technical Report* (OECD, forthcoming).

The survey was implemented through standardised procedures. The PISA Consortium provided comprehensive manuals that explained the implementation of the survey, including precise instructions for the work of School Co-ordinators and scripts for Test Administrators to use during the assessment sessions. Proposed adaptations to survey procedures, or proposed modifications to the assessment session script, were submitted to the PISA Consortium for approval prior to verification. The PISA Consortium then verified the national translation and adaptation of these manuals.

To establish the credibility of PISA as valid and unbiased and to encourage uniformity in administering the assessment sessions, Test Administrators in participating countries were selected using the following criteria: it was required that the Test Administrator not be the mathematics, reading or science instructor of any students in the sessions he or she would administer for PISA; it was recommended that the Test Administrator not be a member of the staff of any school where he or she would administer for PISA; and it was considered preferable that the Test Administrator not be a member of the staff of any school in the PISA sample. Participating countries organised an in-person training session for Test Administrators.

Participating countries and economies were required to ensure that: Test Administrators worked with the School Co-ordinator to prepare the assessment session, including updating student tracking forms and identifying excluded students; no extra time was given for the cognitive items (while it was permissible to give extra time for the student questionnaire); no instrument was administered before the two one-hour parts of the cognitive session; Test Administrators recorded the student participation status on the student tracking forms and filled in a Session Report Form; no cognitive instrument was permitted to be photocopied; no cognitive instrument could be viewed by school staff before the assessment session; and Test Administrators returned the material to the national centre immediately after the assessment sessions.

National Project Managers were encouraged to organise a follow-up session when more than 15% of the PISA sample was not able to attend the original assessment session.

National Quality Monitors from the PISA Consortium visited all national centres to review data-collection procedures. Finally, School Quality Monitors from the PISA Consortium visited a sample of seven schools during the assessment. For further information on the field operations, see the *PISA 2012 Technical Report* (OECD, forthcoming).

Marking procedures were designed to ensure consistent and accurate application of the marking guides outlined in the PISA Operations Manuals. National Project Managers were required to submit proposed modifications to these procedures to the Consortium for approval. Reliability studies to analyse the consistency of marking were implemented.

Software specially designed for PISA facilitated data entry, detected common errors during data entry, and facilitated the process of data cleaning. Training sessions familiarised National Project Managers with these procedures.

For a description of the quality assurance procedures applied in PISA and in the results, see the *PISA 2012 Technical Report* (OECD, forthcoming).

The results of adjudication showed that the PISA Technical Standards were fully met in all countries and economies that participated in PISA 2012, with the exception of Albania. Albania submitted parental occupation data that were incomplete and appeared inaccurate, since there was over-use of a narrow range of occupations. It was not possible to resolve these issues during the course of data cleaning, and as a result neither parental occupation data nor any indices which depend on this data are included in the international dataset. Results for Albania are omitted from any analyses which depend on these indices.

ANNEX A5

THE PROBLEM-SOLVING ASSESSMENT DESIGN

How the PISA 2012 assessments of problem-solving was designed

The development of the PISA 2012 problem-solving tasks was co-ordinated by an international consortium of educational research institutions contracted by the OECD, under the guidance of a group of problem-solving experts from participating countries (members of the problem solving expert group are listed in Annex C of this Volume). Participating countries contributed stimulus material and questions, which were reviewed, tried out and refined iteratively over the three years leading up to the administration of the assessment in 2012. The development process involved provisions for several rounds of commentary from participating countries, as well as small-scale piloting and a formal field trial in which samples of 15-year-olds (about 1 000 students) from participating countries took part. The problem-solving expert group recommended the final selection of tasks, which included material submitted by participating countries. The selection was made with regard to both their technical quality, assessed on the basis of their performance in the field trial, and their cultural appropriateness and interest level for 15-year-olds, as judged by the participating countries. Another essential criterion for selecting the set of material as a whole was its fit to the framework described in Chapter 1 of this volume, in order to maintain the balance across various aspect categories. Finally, it was carefully ensured that the set of questions covered a range of difficulty, allowing good measurement and description of the problem-solving competence of all 15-year-old students, from the least proficient to the highly able.

Forty-two problem-solving questions arranged in 16 units were used in PISA 2012, but each student in the sample only saw a fraction of the total pool because different sets of questions were given to different students. The problem-solving questions selected for inclusion in PISA 2012 were organised into four 20-minute clusters. In countries that also assessed mathematics and reading on computers, computer-based mathematics and digital reading questions were similarly arranged in 20-minute clusters, and assembled together with problem-solving clusters to form test forms (the computer equivalent of paper booklets). In all cases, the total time allocated to computer-based tests was 40 minutes.

In countries that assessed only problem-solving on computers, the four clusters of problem-solving units (CP1-CP4) were rotated so that each cluster appeared twice in each of the two possible positions in the form and every cluster formed two pairs with two other clusters. Eight test forms were built according to the scheme illustrated in Figure A5.1: According to this scheme, each problem-solving item was administered to about one half of all students assessed in problem solving (see Table A2.4).

In those countries that assessed problem solving, mathematics and reading on computers, the four clusters of problem-solving units, the four clusters of mathematics units (CM1-CM4) and the two clusters of reading units (CR1, CR2) were combined into 24 test forms as illustrated in Figure A5.2. One form was chosen at random for administration to each student.

■ Figure A5.1 ■

**PISA 2012 computer-based test design:
Problem solving only**

Form ID	Cluster	
31	CP1	CP2
32	CP2	CP3
33	CP3	CP4
34	CP4	CP1
35	CP2	CP1
36	CP3	CP2
37	CP4	CP3
38	CP1	CP4

■ Figure A5.2 ■

**PISA 2012 computer-based test design:
Problem solving, mathematics and reading**

Form ID	Cluster	
41	CP1	CP2
42	CR1	CR2
43	CM3	CM4
44	CP3	CR1
45	CR2	CM2
46	CM1	CP4
47	CR2	CR1
48	CM2	CM1
49	CP3	CP4
50	CM4	CR2
51	CP1	CM3
52	CR1	CP2
53	CM1	CM3
54	CP4	CP1
55	CR1	CR2
56	CP2	CM4
57	CR2	CP3
58	CM2	CR1
59	CP2	CP3
60	CM4	CM2
61	CR2	CR1
62	CM3	CP1
63	CR1	CM1
64	CP4	CR2



This scheme ensured that every cluster appeared twice in each position for problem solving and computer-based mathematics and four times for digital reading. Moreover, every cluster appeared twice with clusters from a different domain – once in the first and once in the second position within the form. Each of the three domains got the same number of appearances within the 24 forms and therefore an equal proportion of the student sample was assessed in each domain. According to this scheme, each problem-solving item was administered to about one third of all students assessed in problem solving (see Table A2.4), or one sixth of all students assessed on computer.

This design made it possible to construct a single scale of problem-solving proficiency, in which each question is associated with a particular point on the scale that indicates its difficulty, whereby each student's performance is associated with a particular point on the same scale that indicates his or her estimated proficiency. A description of the modelling technique used to construct this scale can be found in the *PISA 2012 Technical Report* (OECD, forthcoming).

References

OECD (forthcoming), *PISA 2012 Technical Report*, PISA, OECD Publishing.



ANNEX A6

TECHNICAL NOTE ON BRAZIL

In 2006, the education system in Brazil was revised to include one more year at the beginning of primary school, with the compulsory school age being lowered from seven to six years old. This change has been implemented in stages and will be completed in 2016. At the time the PISA 2012 survey took place, many of the 15-year-olds in Grade 7 had started their education under the previous system. They were therefore equivalent to Grade 6 students in the previous system. Since students below Grade 7 are not eligible for participation in PISA, the Grade 7 students in the sample were not included in the database.

Brazil also has many rural “multigrade” schools where it is difficult to identify the exact grade of each student, so not possible to identify students who are at least in Grade 7. The results for Brazil have therefore been analysed both with and without these rural schools. The results reported in the main chapters of this report are those of the Brazilian sample without the rural schools, while this annex gives the results for Brazil with the rural schools included.

[Part 1/1]
Table A6.1 **Percentage of Brazilian students at each proficiency level on the problem-solving scale**

		Percentage of students at each level													
		Below Level 1 (below 358.49 score points)		Level 1 (from 358.49 to less than 423.42 score points)		Level 2 (from 423.42 to less than 488.35 score points)		Level 3 (from 488.35 to less than 553.28 score points)		Level 4 (from 553.28 to less than 618.21 score points)		Level 5 (from 618.21 to less than 683.14 score points)		Level 6 (above 683.14 score points)	
		%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.
Problem-solving scale	All	23.5	(1.6)	25.5	(1.4)	26.1	(1.3)	16.8	(1.4)	6.3	(0.8)	1.4	(0.3)	0.4	(0.1)
	Boys	20.8	(1.8)	23.8	(1.5)	25.9	(1.5)	18.3	(1.7)	8.5	(1.2)	2.0	(0.4)	0.6	(0.3)
	Girls	26.0	(1.9)	27.1	(1.9)	26.2	(1.5)	15.3	(1.7)	4.3	(0.7)	0.9	(0.3)	0.1	(0.1)

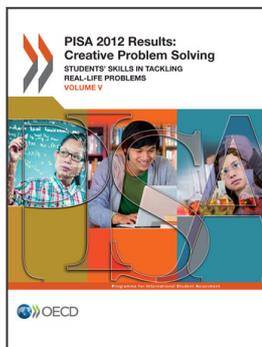
StatLink <http://dx.doi.org/10.1787/888933003744>

[Part 1/1]
Table A6.2 **Mean score, variation and gender differences in student performance in Brazil**

		All students				Gender differences				Percentiles															
		Mean score		Standard deviation		Boys		Girls		Difference (B - G)		5th		10th		25th		50th (median)		75th		90th		95th	
		Mean	S.E.	S.D.	S.E.	Mean score	S.E.	Mean score	S.E.	Score dif.	S.E.	Score	S.E.	Score	S.E.	Score	S.E.	Score	S.E.	Score	S.E.	Score	S.E.		
Problem-solving scale		425	(4.5)	92	(2.3)	436	(5.2)	415	(4.4)	21	(3.3)	273	(5.8)	307	(4.7)	363	(4.8)	426	(5.2)	487	(6.1)	543	(5.7)	573	(5.7)

Note: Values that are statistically significant are indicated in bold (see Annex A3).

StatLink <http://dx.doi.org/10.1787/888933003744>



From:
**PISA 2012 Results: Creative Problem Solving
(Volume V)**
Students' Skills in Tackling Real-Life Problems

Access the complete publication at:
<https://doi.org/10.1787/9789264208070-en>

Please cite this chapter as:

OECD (2014), "PISA 2012 technical background", in *PISA 2012 Results: Creative Problem Solving (Volume V): Students' Skills in Tackling Real-Life Problems*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264208070-11-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.