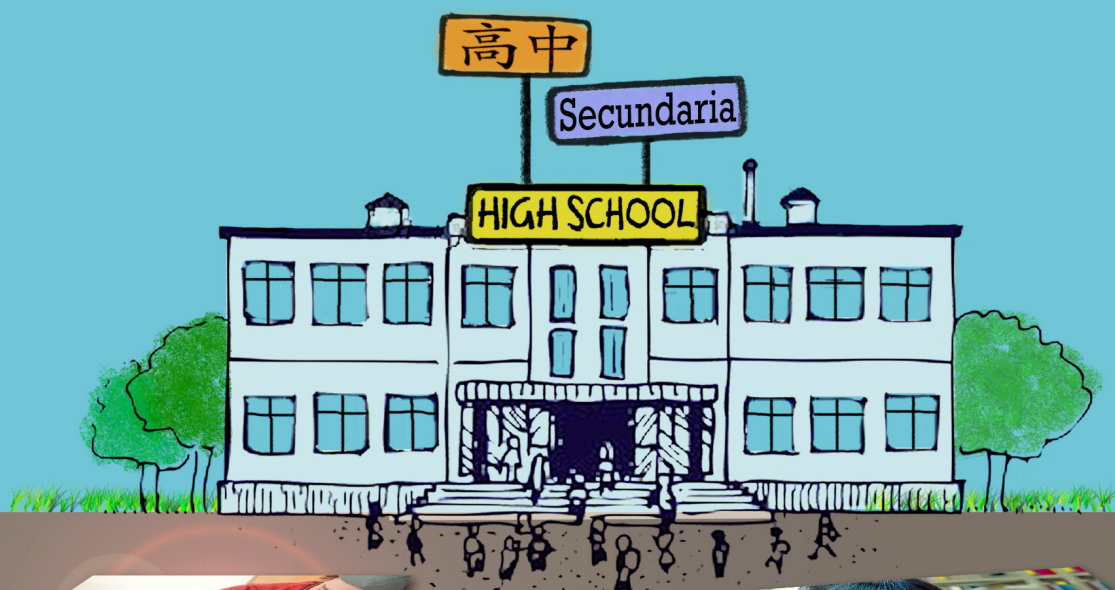




# PISA-BASED TEST FOR SCHOOLS

## TECHNICAL REPORT 2016





PISA-BASED TEST FOR SCHOOLS

# **TECHNICAL REPORT 2016**



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.

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.

Photo credits:

© Khoa Vu/Flickr/Getty Images  
© Shutterstock/Kzenon  
© Simon Jarratt/Corbis  
© Stocklib Image Bank/No. 13883137

© OECD 2017

---

This work is available under the [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO](https://creativecommons.org/licenses/by-nc-sa/3.0/) (CC BY-NC-SA 3.0 IGO). For specific information regarding the scope and terms of the licence as well as possible commercial use of this work or the use of PISA data please consult [Terms and Conditions](https://www.oecd.org/termsandconditions/) on [www.oecd.org](https://www.oecd.org).

---

## *Foreword*

The OECD PISA-based Test for Schools assessment is geared for use by schools and networks of schools around the world to support research, international benchmarking and school improvement efforts. It collects information about 15-year-old students' applied knowledge and competencies in reading, mathematics and science as well as their attitudes toward learning and school. PISA-based Test for Schools examines how students in the participating schools are prepared to meet the challenges of the future. The data collected by the assessment are an extremely valuable source of information for school principals, educators, parents and students.

The methodology of the PISA-based Test for Schools is complex and demanding. The PISA-based Test for Schools Report describes those procedures and methodologies along with other features that enable the PISA-based Test for Schools to provide high-quality data to schools and local school administrations wanting to go further in understanding how their own individual schools perform compared with the world's leading school systems. The descriptions are provided at a level that will enable replication of all implementation procedures.

This *Technical Report* concerns the paper-and-pencil version of the PISA-based Test for Schools.

The first edition of this report was drafted by Noémie Le Donné, Tue Halgreen and Kelly Makowiecki. The 2016 edition of the report was prepared by Javier Suárez-Álvarez with advice from Francesco Avvisati, Isabel Benitez, Richard Li, François Seyler and Tse Chi Sum, and under the supervision of Tue Halgreen. Administrative support for the 2016 edition was provided by Florence Bernard and Sophie Limoges co-ordinated production.



## *Table of Contents*

1. Introduction .....	9
a. Features of the PISA-based Test for Schools.....	9
b. Managing and implementing the PFS: Overview of roles and responsibilities .....	10
c. Structure of this report .....	11
d. Additional information.....	12
<b>Chapter 1 Instrument design</b> .....	13
2. Test design, scope and development .....	13
a. Test design .....	13
b. Test scope.....	14
c. Test development .....	15
d. Reporting PFS results on PISA scales .....	16
3. Context Questionnaires.....	17
a. Student Questionnaire .....	17
b. School Questionnaire .....	19
<b>References</b> .....	21
<b>Chapter 2 Operations</b> .....	23
4. Overview of the roles and responsibilities.....	23
a. School Coordinators.....	23
b. Test Administrators.....	24
5. Preparing for implementation .....	25
a. Translation, adaptation and verification of the test and survey material.....	25
b. Field trial participation and outcomes.....	27
6. Field operations .....	28
a. School and student eligibility to participate in the survey .....	28
b. Sampling .....	29
c. Packaging and shipping materials.....	32
d. Receipt of materials by the National Service Provider after testing .....	32
e. Coding of the cognitive test and of the context questionnaires.....	33
f. Coding the test booklets.....	33
g. Student questionnaire coding .....	34
h. School questionnaire coding .....	35
<b>Chapter 3 Data Processing</b> .....	37
7. Constructing the initial PFS database .....	37
a. Files in the initial PFS database .....	37
b. Records in the database .....	39
c. Representing missing data .....	40
d. Merging the data files .....	41
8. Procedures for scaling cognitive data .....	41

a. The mixed coefficients multinomial logit model .....	41
b. Application to PFS .....	44
9. Procedures for constructing contextual variables .....	52
a. Overview .....	52
b. Simple questionnaire variables .....	53
c. Questionnaire scale indices .....	54
d. Other complex questionnaire variables .....	62
10. Survey weighting (the balanced repeated replication method) .....	80
a. The balanced repeated replication method .....	80
b. The sampling variance estimator .....	83
11. Statistical procedures for generating school report results .....	83
a. Univariate statistics for context variables .....	83
b. Students average scores and their standard errors .....	85
c. Reporting and reviewing school-level data .....	88
References .....	89
<b>Chapter 4 School Report</b> .....	93
12. Supporting the preparation of a national version of the school report template .....	93
a. Proposal for adaptations .....	94
b. Translation of the template .....	94
13. Production of the school report eBooks .....	94
a. Automatizing the school report production .....	94
b. Receiving OECD's approval on the eBooks .....	95
14. Grouped school report .....	96
<b>Annex 1 Technical documentation</b> .....	97
<b>Annex 2 Examples of R codes for scaling</b> .....	100
 <b>Tables</b>	
Table 1. Cluster rotation design used to form test booklets for PFS .....	13
Table 2. Item types .....	15
Table 3. Number of items .....	16
Table 4. Lower score limits for the proficiency levels in reading, mathematics and science .....	52
Table 5. OECD means and standard deviations of Weighted Likelihood Estimates (WLEs) .....	56
Table 6. Item parameters for mathematics self-efficacy (MATHEFF) .....	56
Table 7. Item parameters for instrumental motivation for mathematics (INSTMOT) .....	59
Table 8. Item parameters for science self-efficacy (SCIEEFF) .....	59
Table 9. Item parameters for instrumental motivation to learn science (INSTSCIE) .....	60
Table 10. Item parameters for disciplinary climate in language of instruction (DISCLIMA) .....	61
Table 11. Item parameters for disciplinary climate in mathematics (DISCLIM, PISA 2012) .....	61
Table 12. Item parameters for disciplinary climate in science (DISCLICI, PISA 2015) .....	62
Table 13. Item parameters for teachers student relations (STUDREL) .....	62
Table 14. Factor loadings for PISA 2015 participating countries and economies .....	64
Table 15. Mapping of ISCED to years .....	67
Table 16. Home possession items .....	72
Table 17. International item parameters for HOMEPOS .....	73



Table 18. National item parameters for HOMEPOS .....	75
Table 19. Hadamard Matrix .....	82
Table 20. Replicates for the Balanced Replicate method .....	82
Table 21. The 405 mean estimates .....	85
Table 22. Link error estimates .....	87

## Figures

Figure 1. Example of scatter plot .....	47
Figure 2. Computation of ESCS in PISA 2015 .....	63
Figure 3. How the reading process can be characterised .....	77
Figure 4. Profile of readers .....	78
Figure 5. Latent profile analysis with two categorical latent variables .....	80

## Boxes

Box 1. Examples of participating schools eligible to receive school report .....	40
Box 2. ConQuest® syntax for running the item analyses on the reading test .....	45
Box 3. ConQuest® syntax for generating the plausible values in reading .....	51
Box 4. ConQuest® syntax for generating the WLEs of the MATHEFF scale .....	57
Box 5. Contents of the item parameter file matheff.prm .....	58
Box 6. First six lines of the WLE file matheff.wle .....	58
Box 7. Scaling procedures of Context Questionnaire in PISA 2015 .....	70



## 1. Introduction

### *a. Features of the PISA-based Test for Schools*

The PISA-based Test for Schools (PFS) is a student assessment tool geared for use by schools and networks of schools to support research, international benchmarking and school improvement efforts. In the United States, the assessment is known as the OECD Test for Schools (based on PISA). The assessment tool provides descriptive information and analyses on the skills and creative application of knowledge of 15-year-old students in reading, mathematics, and science, comparable to existing PISA scales.

The assessment also provides information on how different factors within and outside school associate with student performance. Contextual questionnaires geared for students and schools are an important part of the assessment. Information about students' socio-economic backgrounds, their attitudes and interests in reading, science and mathematics and the learning environment at school are all addressed in the assessment.

The PFS provides important peer-to-peer learning opportunities for educators – locally, nationally and internationally – as well as the opportunity to share good practices to help identify “what works” to improve learning and build better skills for better lives.

The OECD completed the international pilot trial of the assessment in March 2013. Since 2010 and under the guidance of the PISA Governing Board (PGB), the OECD has carried out the development of the assessment and the implementation of the pilot in collaboration with schools and local partners in different countries.

The organisational and technical characteristics of the PFS involve a number of different aspects:

- The design of the test and the features incorporated into the test developed for PFS.
- The student sampling design, requirement and procedures.
- Because of the international nature of the test, rules and procedures required to guarantee the equivalence of the different language and cultural versions used within and between participating countries.
- Various operational procedures, including test administration arrangements, data capture and processing, and quality assurance mechanisms designed to ensure the generation of comparable data from all schools and countries.
- Scaling and analysis of the data and their subsequent reporting: PFS employs scaling models based on Item Response Theory (IRT) methodologies. The described proficiency scales, which are the basic tool in reporting PFS outcomes, are derived using IRT analysis.
- Procedures for automating the production of school reports presenting schools' results at PFS.

This report describes the above-mentioned methodologies, as they must be implemented in the PFS by the National Service Provider.

***b. Managing and implementing the PFS: Overview of roles and responsibilities***

The PFS is coordinated in each country by an accredited National Service Provider who implements the procedures that were prepared and agreed upon with the OECD. This report has been developed for National Service Providers to acquaint themselves with the PFS design and with the procedures required to correctly implement the PFS, and generate and report back results to the participating schools.

The National Service Provider is responsible for implementing the project within its own country. It:

- is in contact with the OECD and receives support in all aspects of the PFS operational procedures;
- discusses nationally specific aspects of the implementation of the PFS (such as national options regarding sampling, analyses and reporting);
- establishes procedures for the security and protection of the confidentiality of materials during all phases of the implementation;
- prepares the national versions of the test instruments, questionnaires, manuals and coding guides;
- conducts a field trial and analyses the field trial data in order to assess and possibly improve the quality of the survey instruments;
- recruits the participating schools;
- identifies school coordinators from each of the participating schools (nominated by the school principal or a volunteer from the school staff) and works with them on school preparation activities;
- selects the student sample from a list of eligible students provided by the school coordinators;
- recruits and trains tests administrators according to the PFS technical standards to administer the tests within schools;
- recruits and train coders;
- arranges for the data entry of the test and questionnaire responses;
- processes school cognitive and context data according to the procedures described in this *Technical Report* and generating schools results; and
- coordinates the reporting of individual school results and sends school reports back to the participating schools.

If the National Service Provider does not have all the capacities required for carrying out each of the above-listed tasks, it can outsource some of the responsibilities to one or several other service providers accredited by the OECD.

The team of the National Service Provider will typically consist of:

- a Project Manager who will oversee and manage the project, and liaise with the OECD, the schools and other key partners of the project;
- a team of Test Administrators who will be responsible for administering the PFS to the schools in accordance with PFS standard procedures;
- a team of Coders who will score students' responses to the cognitive test and the Student Questionnaire and schools' responses to the School Questionnaire;
- a Statistician(s), preferably with experience with IRT modelling, who will process the data and generate the results;
- a Layout/Graphic Designer(s) who will prepare school reports based on a report template prepared by the OECD.

The staff resources will of course depend on the local setup and the number of schools expected to participate in the test.

### *c. Structure of this report*

This *Technical Report* is designed to describe organisational and technical aspects of the project at a sufficient level of detail to enable replication of the procedures for managing and implementing the PISA-based Test for Schools assessment. There are four parts and an annex in the report:

- Part 1 – Instrument design: describes the design and development of both the cognitive test and the contextual questionnaires (chapters 2 and 3).
- Part 2 – Operations: gives details of the operational procedures for instrument translation, the sampling and population definitions, and test administration (chapters 4 to 6).
- Part 3 – Data processing: covers the procedures used for constructing the database, performance and context variables, including scaling methods, analysing the field trial data, and generating school report results (chapters 7 to 11).
- Part 4 – School report: covers the procedures used for supporting the preparation of the school report template, and producing the school report eBooks (chapters 12 and 13).
- Annex: provides a list of the technical materials provided to the National Service Provider by the OECD.

*d. Additional information*

Any additional technical issues not addressed in this report, such as specific requests from participating schools, must first be discussed with the OECD.

## Chapter 1

### Instrument design

#### 2. Test design, scope and development

This section describes the test design for the PISA-based Test for Schools and the processes by which the Australian Council for Educational Research (ACER) developed the cognitive tests for mathematics, reading and science and linked them to the PISA tests.

##### *a. Test design*

The test items for mathematics, reading and science are compiled in seven item clusters (two mathematics clusters, two reading clusters, two science clusters and one cluster including items from all three domains) with each cluster representing 40 minutes of test time. The items are presented to students in seven standard test booklets, with each booklet being composed of three clusters, hence two hours of test time. Clusters labelled M1 and M2 denote the two mathematics clusters, R1 and R2 denote the reading clusters, S1 and S2 denote the science clusters, and RMS denotes the cluster with items from all three domains.

The cluster rotation design for the booklets are similar to designs used in PISA surveys and is shown in Table 1.

**Table 1. Cluster rotation design used to form test booklets for PFS**

Booklet ID	Cluster 1	Cluster 2	Cluster 3
1	R1	RSM	M1
2	RSM	M2	S2
3	M2	M1	R2
4	M1	S2	S1
5	S2	R2	R1
6	R2	S1	RSM
7	S1	R1	M2

This is a balanced incomplete block design. Each cluster (and therefore each test item) appears in three of the three-cluster test booklets, once in each of the three possible positions within a booklet, and each pair of clusters appears in one (and only one) booklet.

Each sampled student is randomly assigned to one of the seven booklets administered in each school, which means each student undertakes two hours of testing.

### ***b. Test scope***

PFS items are arranged in units based around a common stimulus. Many different types of stimulus are used including passages of text, tables, graphs and diagrams, often in combination. Each unit contains from one to five items assessing students' competencies and knowledge. A complete PFS unit consists of some stimulus material, one or more items (questions), and a guide to the coding of responses to each question. Each coding guide comprises a list of response categories (full, partial and no credit), each with its own scoring code, descriptions of the kinds of responses to be assigned each code, and sample responses for each response category. Coding guides for each domain of assessment will be provided by the OECD.

The mathematics assessment consists of 25 units comprising a total of 40 cognitive items, the reading assessment consists of 47 items (13 units), the science assessment consists of 54 items (25 units), each representing about 130 minutes of testing time.

Item formats are either selected response or constructed response. Selected response items are either simple multiple-choice with several responses from which students are required to select the best answer, or complex multiple-choice presenting several statements for each of which students are required to choose one of two or more possible responses (yes/no, true/false, correct/incorrect, etc.). Constructed response items are of two broad types. Constructed response manual items require limited manual input by trained coders at the stage of processing student responses. They require students to construct a numeric response within very limited constraints, or only required a word or short phrase as the answer, and coders later to assign each response to the predefined response categories. Constructed response expert items require the use of trained expert coders to interpret observed student responses and assign them to one of the defined response categories. These items require a response to be generated by the student, with a range of possible full-credit answers. Note that although there are a total of 141 items across Mathematics, Reading and Science, one reading item, PR6004Q05A, is not scored separately. The response to this question is taken into consideration in the scoring of item PR6004Q05B. Table 2 shows the number of items of each type.



**Table 2. Item types**

	Total number of items	Number of items				
		Constructed response expert items	Constructed response manual items	Complex multiple choice items	Simple multiple choice items	Not scored items
Reading	47	17	4	7	18	1
Mathematics	40	7	19	3	11	0
Science	54	20	0	16	18	0
<b>Total</b>	141	44	23	26	47	1

Pencils, erasers, rulers, and in some cases calculators, must be provided to students undertaking the PFS assessment. It is recommended that calculators be provided in schools for a country where they are routinely used in the classroom. National Service Providers can decide whether calculators should be provided for the students on the basis of standard national practice. No test item requires a calculator, but some mathematics items involve solution steps for which the use of a calculator could be of assistance to students accustomed to their use.

### *c. Test development*

Experience gained in former OECD assessments, such as PISA, showed the importance of collaborating with an experimented test centre to help achieve conceptually rigorous material that has the highest possible levels of cross-cultural and cross-national diversity. Accordingly, all item development was undertaken at ACER which was responsible for the item development of the first four PISA waves. Test development for the PFS survey commenced in 2010. Development proceeded through various processes and stages, slightly different for each of the cognitive domains in which test material was required, and culminating in 2012 in the selection of items proposed for use in the main survey. This section presents the development arrangements and approaches taken by ACER to produce the material required.

The material needed to fulfil the design requirements had to satisfy the domain definitions and specifications within the relevant assessment framework. For each subject domain, the PISA assessment framework was used to develop the PFS assessment. The framework defines the domain, describes the scope of the assessment, specifies the structure of the test – including item format and the preferred distribution of items according to important framework variables – and outlines the possibilities for reporting results. Detail on the PISA assessment framework for reading, mathematics and science are published in *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy* (OECD, 2013).

The test development teams of ACER conducted development of items, including cognitive laboratory activities, in English. Each domain team included individuals who have been involved in the test development for the main PISA surveys.

A total of 420 cognitive items were developed by ACER in two phases. All items were field trial tested, along with 63 PISA link items, across students from schools in Australia, Ireland, the United Kingdom and the United States. Data from the international field trial was analysed using standard item response techniques.

The 420 available items were evaluated by the expert group in terms of their substantive quality, fit to framework, range of difficulty, psychometric quality, durability and interest level for 15-year-olds.

The selection of items to be proposed for inclusion in the main survey instruments had to satisfy the following conditions:

- The psychometric properties of all selected items had to be satisfactory.
- There had to be an appropriate distribution of item difficulties, broad enough to generate useful measurement data at both extremes of the anticipated ability distribution of sampled students across all participating countries.

Characteristics of the item set used in the field trial, and the selected set for the main survey, are presented in the *PISA-based Test for Schools: Technical Report* produced by ACER (2012).

In selecting PISA link items, framework balance, range of difficulty, and a high level of reliability were considered as prime criteria.

Table 3 gives the original number of items, the number of selected items and the number of PISA link items, per domain.

**Table 3. Number of items**

Domain	Original Pool	Selected Pool	PISA-linked items
Reading	100	47	27
Mathematics	75	40	16
Science	114	54	20
<b>Total</b>	289	141	63

#### ***d. Reporting PFS results on PISA scales***

The PISA scale for reading was developed in PISA 2000, when reading was the major domain of assessment. The scale was established so that the mean and standard deviation of the PISA 2000 scores was 500 and 100 respectively, for the equally weighted 27 OECD countries that participated in PISA 2000 that had acceptable response rates (Adams and Wu, 2002). For PISA 2003, 2006, 2009, 2012 and 2015 the decision was made to report the reading scores on this previously developed scale, so the reading reporting scales used for PISA 2000, 2003, 2006, 2009, 2012 and 2015 are directly

comparable. The value of 500, for example, has the same meaning in any of these evaluations as it did in PISA 2000.

Mathematics, was the subject of major development work for PISA 2003. For mathematics the reporting scale was determined such that the mean was 500 and standard deviation 100 for the 30 OECD countries that participated in PISA 2003. For PISA 2006, 2009, 2012 and 2015 the decision was made to report the mathematics on the PISA 2003 scale.

For science a new scale was established in 2006. The metric for that scale was set so that the mean was 500 and standard deviation 100 for the 30 OECD countries that participated in PISA 2006. For PISA 2009, 2012 and 2015 the decision was made to report the science scores on the PISA 2006 scale.

To permit a comparison of the PFS results with those of PISA, the decision was thus made to report:

- the PFS reading scores on the PISA reading scale developed for PISA 2000;
- the PFS mathematics scores on the PISA mathematics scale developed for PISA 2003; and
- the PFS science scores on the PISA science scale developed for PISA 2006.

Further details on the various PISA reporting scales are given in Chapters 9 and 12 of the *PISA 2012 Technical Report* (OECD, 2014).

The linking of PFS mathematics, reading and science to the existing PISA scales was undertaken using equating methods. The mathematics and science items were put on the current PISA scale (2003 and 2006 respectively) by anchoring the PISA-linked items at their 2003 and 2006 item parameters. The reading items were mapped on to the 2000 scale by adjusting the PISA 2009 item parameters by a linear transformation to the PISA 2000 scale. Details are provided in the *PISA-based Test for Schools: Technical Report* produced by ACER (2012). The methodology of PFS was revised and improved after the first testing cycle. The improved methodology applies to all participating countries from November 2015, with details provided in the following technical note: <https://www.oecd.org/pisa/aboutpisa/Technical-note-revised-methodology.pdf>. The transformations for putting the PFS logit scores on the PISA point scales are given in Section 8.b.ii of Part 3.

### 3. Context Questionnaires

This section describes the content of the Student and School Questionnaires to be used as part of the PISA-based Test for Schools.

#### a. Student Questionnaire

The PFS Student Questionnaire includes questions from five PISA Student Questionnaires: PISA 2003, 2006, 2009, 2012 and 2015. It has the same general structure

as the PISA Student Questionnaires. Questions were selected for inclusion for their comparability with recent PISA results and for their analytical relevance for school-level reporting. The first part of the questionnaire includes core questions on the student's family background and school experience and the second part includes subject-specific questions. The subject-specific questions are ordered according to subjects with two sections of reading questions, two sections of science questions and one section of mathematics questions. The section on reading activities outside school is placed before the questions related to the school, similar to the structure of the main PISA 2009 Student Questionnaire.

The structure with section headings is:

- SECTION 1: ABOUT YOU
- SECTION 2: YOUR FAMILY AND YOUR HOME
- SECTION 3: YOUR SCHOOL
- SECTION 4: YOUR LEARNING TIME
- SECTION 5: YOUR READING ACTIVITIES AND STRATEGIES
- SECTION 6: YOUR MATHEMATICS LEARNING EXPERIENCE
- SECTION 7: YOUR <TEST LANGUAGE> LEARNING EXPERIENCE
- SECTION 8: YOUR <BROAD SCIENCE> LEARNING EXPERIENCE

Questionnaire items are organised into units. Units consist of multiple-choice questions presenting one or several statements for each of which students are required to choose the best answer(s), or one or several constructed response questions.

A questionnaire unit is identified by a short label. The units' labels consist of seven characters. The first two characters denote the PISA Source Questionnaire and are 09, 12 and 15, respectively. They are always followed by the three characters "-ST". The next two characters indicate the unit within the PISA Source Questionnaire. For example, 09-ST01 is the first unit of the PISA 2009 Student Questionnaire. In the PFS database, it is recommended to form item names by adding an item number to the unit label. For example, 09-ST08Q01 will indicate the student response to whether the student's mother lives at home with the student, 09-ST08Q02 the student response to whether the student's father lives at home with the student, etc.

The following questions in the PFS Student Questionnaire are constructed response manual items that require very limited manual input by trained coders at the stage of processing student responses:

15-ST01: What year group are you in?

15-ST021: If you were not born in <test country>, how old were you when you arrived in <test country>?

Two sets of questions in the PFS Student Questionnaire are constructed response (expert) items that require the use of trained coders to interpret observed student responses and assign them to one of the defined response categories:

15-ST014: What is your mother’s main job? What does your mother do in her main job?

15-ST0158: What is your father’s main job? What does your father do in his main job?

Details on how to code student responses to these questions are provided in Section 6.e of Part 2.

The total estimated response time to the PFS Student Questionnaire for each student is approximately 23 to 27 minutes.

### ***b. School Questionnaire***

There are two existing versions of the School Questionnaire. The reduced version of the School Questionnaire is designed to only collect information required for generating results reported back to participating schools. The extended version of the School Questionnaire was designed and administered for the PFS international pilot study. The anonymous data gathered through this questionnaire will be used for research purposes on school management. National Service Providers, in consultation with the OECD, can choose to administer either version of the School Questionnaire.

#### *i. Reduced version of the School Questionnaire*

The reduced version of the School Questionnaire is designed to collect information required for reporting purposes only. It includes questions about:

- the total number of students enrolled in the school;
- the school type and sector; and
- the school location.

The estimated response time to the reduced School Questionnaire is a couple of minutes.

#### *ii. Extended version of the School Questionnaire*

The extended version of the PFS School Questionnaire includes all questions from the PISA 2009 School Questionnaire that were retained in PISA 2012, which allows for comparability between the PFS and most recent PISA School Questionnaires. Questions that were used to construct school-level indices and results in PISA 2009 initial reports are also included. As in the main PISA assessment, it covers such elements as:

- the structure and organisation of the school;

- student and teachers demographics;
- the school's resources;
- the school's instruction, curriculum and assessment;
- the school climate; and
- school policies and practices.

The total estimated response time of this extended version of the School Questionnaire is around 30 minutes, similar to the School Questionnaire in the main PISA assessment.

## References

- ACER (2012), *PISA-based Test for Schools: Technical Report*, Australian Council for Educational Research (ACER).
- Adams, R.J. and M.L. Wu (2002), *Programme for International Student Assessment (PISA): PISA 2000 Technical Report*, PISA, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264199521-en>.
- OECD (2014), *PISA 2012 Technical Report*, OECD, Paris, <http://www.oecd.org/pisa/pisa-products/PISA-2012-technical-report-final.pdf>.
- OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, PISA, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264190511-en>.
- OECD (2012), *PISA 2009 Technical Report*, OECD, Paris, <http://www.oecd.org/pisa/pisa-products/50036771.pdf>.
- OECD (2009), *PISA 2006 Technical Report*, OECD, Paris, <http://www.oecd.org/pisa/pisa-products/42025182.pdf>.





## *Chapter 2*

## Operations

### 4. Overview of the roles and responsibilities

For school-level operations, the National Service Provider co-ordinates activities with school-level staff, referred to as School Coordinators. Trained Test Administrators administer the PFS in schools.

#### *a. School Coordinators*

The National Service Provider must identify a School Coordinator in each participating school: typically, this person is nominated by the school principal or is a volunteer from the school staff. School Coordinators coordinate school-related activities with the National Service Provider and the Test Administrators.

The School Coordinators:

- Establish the testing date and time in consultation with the accredited National Service Provider.
- Prepare a student list with names of all eligible students in the school and send it to the National Service Provider so that the latter can select the student sample.
- Receive the list of sampled students on a student tracking form from the National Service Provider and update it if necessary, including identifying students with disabilities or limited test language proficiency who cannot take the test according to the criteria established.
- Administer the school questionnaire.
- Inform school staff, students and parents of the nature of the test and the test date by sending a letter or organising a meeting, and secure parental permission if required by the school.
- Assist the test administrator with room arrangements for the test day.

On the test day, the School Coordinator is expected to ensure that the sampled students attend the test session(s). If necessary, the School coordinator also arranges follow-up session and ensures that absent students attend the follow-up session. The

*School Coordinator's Manual* describes in detail the activities and responsibilities of the School Coordinator.

***b. Test Administrators***

The Test Administrators are primarily responsible for administering the PFS test fairly, impartially and uniformly, in accordance with international standards and main PFS procedures. To maintain fairness, the Test Administrators are usually employed by the National Service Provider. Prior to the test date, Test Administrators need to be trained by the National Service Provider. Training includes a thorough review of the *Test Administrator's Manual*, and the script to be followed during the administration of the test and questionnaire. Additional responsibilities include:

- Ensuring receipt of the testing material from the National Service Provider and maintaining their security.
- Co-operating with the School Coordinator.
- Contacting the School Coordinator one to two weeks prior to the test and confirm plans.
- Completing final arrangements on the test day.
- Conducting a follow-up session, if needed, in consultation with the School Coordinator.
- Reviewing and updating the student tracking form.
- Completing the session attendance form (a form designed to summarise session times, any disturbance to the session, etc.).
- Ensuring that the number of tests and questionnaires collected from students tallies with the number sent to the school.
- Obtaining the school questionnaire from the School Coordinator.
- Sending the school questionnaire, the student questionnaires and all test materials (both completed and not completed) to the National Service Provider after the testing is carried out.

The *Test Administrator's Manual* describes in detail the activities and responsibilities of the Test Administrators.

Templates for the student tracking form and session attendance form will be provided by the OECD.

## 5. Preparing for implementation

### *a. Translation, adaptation and verification of the test and survey material*

The PFS assessment is intended to be available in a large number of countries with different languages, different cultures and different school systems. The PFS follows the main PISA practice of providing the testing materials in the language of instruction in the content areas being tested. The aim is to assess 15-year-old students' skills in three major domains: reading, mathematical and science literacy, using strictly equivalent test batteries, translated and adapted into each of the languages of instruction of the participating schools. In order to achieve this, PFS implements strict procedures for translation/adaptation and verification of assessment materials.

The procedures firstly consisted in the development of a source version of the instruments in English (see Section 0 of Part 1). If the language of instruction is not the source language (English), a full translation, as described here, is needed. For English-language and non-English language applications, some degree of adaptation of the materials to local educational and cultural terminology is needed.

#### *i. Language of instrumentation and administration*

In some countries, the PFS is to be administered in more than one language, so the procedures for translation and adaptation need to be carried out for each language. The decision about choice of the test language can be made at the student, school or country level. These situations require agreement with the OECD.

#### *ii. Materials to be prepared*

The key end-user materials need to be translated into the assessment language or languages so as to be linguistically equivalent to the PFS source version:

- All administered assessment items
- All administered context questionnaires
- Test Administrator manual
- School Coordinator manual
- Coding guides for the cognitive items

#### *iii. Procedures for preparing the materials*

Prior to the field trial (see Section 5.b), the National Service Provider must prepare a translation plan, that clarifies the procedures to be used to develop their national version(s) and the different processes used for translator/reconciler recruitment and training. The translation plan should be reviewed by the OECD for agreement. The translation/adaptation process results in a report prepared by the National Service

Provider, which must be reviewed and approved by the OECD. Test units, questionnaire items, manuals and coding guides are initially sent to National Service Providers several months before the testing dates, allowing adequate time for materials to be translated (if required), adapted and verified. A single translation of the materials should be undertaken by professional translators. The *PFS Translation and Adaptation Guidelines* manual contains general instructions, a number of recommendations to address common translation traps encountered when translating test materials, a list of adaptations that are desirable, acceptable or crucial when translating cognitive tests, notes on translation and adaptation of questionnaires and manuals. Note-versions of the student questionnaire and cognitive items are also available. They contain notes on each question and item and are designed to aid National Service Providers and translators in adapting and translating the survey material in a consistent and internationally comparable way.

### Cognitive test material

National Service Providers are required to submit the translated/adapted items in units, i.e., in sets of items associated with the same stimuli or support, and not in booklets. The cognitive items must be submitted along with a form documenting any proposed national adaptations for verification by the OECD. As in main PISA, one of the most important quality control procedures implemented to ensure high-quality standards in the translated assessment items for PFS is to have an independent team of expert verifiers confirm each national version against the English source version. The OECD establishes one verification centre for being in charge of the linguistic verification of the cognitive items. Verification is not required when a testing language will be used for only less than 500 students across a school year (about 7 schools). Once the verification of the cognitive items has been approved by the OECD, the National Service Provider incorporates the requested modifications into their materials and assembles the seven different test booklets with clusters of test items arranged according to the test booklet design.

Test items are presented in units (stimulus material and items relating to the stimulus) and each cluster contains several units. Units allocated to clusters and clusters allocated to booklets are also provided, together with detailed instructions to the National Service Provider about how to assemble their translated or adapted clusters into booklets. For reference, the source version of all test booklets is provided to National Service Providers in English. National Service Providers are encouraged to use the cover design provided by the OECD. In formatting translated or adapted test booklets, they have to follow the layout in the source version as much as possible, including allocation of items to pages. The seven test booklets are submitted to the OECD, which perform a final check of the materials. This is a verification of the layout, instructions to the student, the rendering of graphic material, etc. Once feedback from OECD's final check has been received and incorporated into the test booklets, the National Service Provider is ready to send the materials to print.

## Context questionnaires

As with the test material, the source versions of the context questionnaires in English are provided to National Service Providers for translation and adaptation into the test languages. As all the questionnaire questions come from the PISA context questionnaires, National Service Providers whose country has already participated in main PISA studies can use the questions translated and adapted for their country (and their language). National Service Providers are permitted to add questions of national interest as national options to the questionnaires. Proposals and text for these are submitted to the OECD for approval as part of the process of reviewing adaptations to the questionnaires. It is recommended that the additional material should be placed at the end of the international modules. With the OECD's approval, National Service Providers can also take out questions that are not relevant for the local context and that are not used in the computation of key school-level results.

National Service Providers are required to submit a note documenting all proposed national adaptations to questionnaire items to the OECD for approval. National Service Providers implement OECD's feedback in the final version of the questionnaires, which is submitted once more in order to conduct a final check. Following feedback from the final check, National Service Providers make final changes to their questionnaires prior to printing.

## School Coordinator and Test Administrator manuals

The *School Coordinator's Manual* and *Test Administrator's Manual* are also required to be translated into the language of instruction. English versions of each manual are provided by the OECD. National Service Providers are required to submit a form documenting all proposed national adaptations to the manuals to the OECD for approval. Following approval of the adaptations, the manuals are prepared and submitted to the OECD once more. National Service Providers implement final OECD feedback into their manuals prior to printing.

All survey instruments that have been translated, adapted and verified are the OECD's property and are made available to new participating countries.

### ***b. Field trial participation and outcomes***

The PFS survey can be made publicly available to schools on demand once a country has successfully implemented a field trial. A field trial is usually only required in an assessment language if that language group represents more than 500 students (i.e., more than 7 schools) expected to be assessed throughout the school year. The sample size for the field trial is a function of the test design and is set to achieve the standard of 200 student responses per cognitive item (i.e., 500 students from a minimum of 7 schools) need to be tested during the field trial. The schools selected for participating in the field trial should be as diverse as possible in terms of level of achievement, school's size, intake and type. The National Service Provider may want to consider increasing the

number of students tested if it wants to deliver reports to the participating schools during the field trial. If so, a minimum of 2 000 observations is required if the model is to be fitted appropriately (see page 44 for further explanation).

During the field trial, the National Service Provider sets up a query service. School Coordinators and Test Administrators are encouraged to send queries to the service so that a common adjudication process is consistently applied to all questions about test administration procedures conducted in the schools. All practical issues that have arisen during the field trial operations, as well as the solutions proposed by the National Service Provider to address and improve them for the publicly available study, must be documented.

The National Service Provider is in charge of analysing the field trial data: all response information from the field trial must be coded, and a validation study of the field trial instruments must be conducted through psychometric analyses of cognitive items. Analyses of cognitive items are to be performed for each language of instruction available in the country (as described in Part 3). Particular attention must be paid to the fit of the items to the scaling model, item discrimination (and item-by-language interactions if there are several instruction languages in the participating country). The results of these analyses are used to assess and possibly improve the quality of the instruments and to make decisions regarding item treatment for each language. This means that an item may be deleted from the PFS scaling in a country (or in a language) if it has poor psychometric characteristics in this particular country (or for this particular language). Depending on the results of the cognitive item analyses, the National Service Provider can be asked by the OECD to perform analyses of scaling outcomes for each language of instruction available in the country (as described in Part 3). Particular attention must then be paid to the tests targeting and reliability, and the domain inter-correlations. The National Service Provider can outsource these analyses to another service provider.

The National Service Provider prepares a report that documents proposed changes in the administration procedures and includes the psychometric analyses. This report must be reviewed by the OECD for approval.

## 6. Field operations

### *a. School and student eligibility to participate in the survey*

#### *i. Target definition*

The National Service Provider is responsible for the recruitment of schools. All schools in the country are eligible to participate in the PFS test if they meet the minimum requirements. The schools must have 35 or more students who are between 15 years and 3 completed months to 16 years and 2 completed months at the time of assessment. The operational definition of an age population directly depends on the testing dates. A variation of up to one month in this age definition is permitted.

### *ii. Recommendations regarding testing time*

Testing is not recommended:

- During the first six weeks of the school year because of a concern that student performance levels may have been lower at the beginning of the academic year than at the end of the previous academic year, even after controlling for age.
- During the assessment period of other OECD tests such as PISA, TALIS, etc.

For cost purposes, it is recommended to offer schools to be tested during a limited period rather than throughout the school year.

### *iii. Within-school exclusions*

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

- Intellectually disabled students are students who have a mental or emotional disability and who, in the professional opinion of qualified staff, are cognitively delayed such that they cannot be validly assessed in the PFS testing setting. This category includes students who are emotionally or mentally unable to follow even the general instructions of the test. Students are not to be excluded solely because of poor academic performance or normal discipline problems.
- Functionally disabled students are students who are permanently physically disabled in such a way that they cannot be validly assessed in the PISA testing setting. Functionally disabled students who can provide responses are to be included in the testing.
- Students with insufficient assessment language experience are students who need to meet all of the following criteria: i) are not native speakers of the assessment language(s); ii) have limited proficiency in the assessment language(s); and iii) have received less than one year of instruction in the assessment language(s). Students with insufficient assessment language experience can be excluded.

### ***b. Sampling***

Student sampling is undertaken using a list of all PFS-eligible students in each school that agrees to participate. The lists can be prepared at the national, regional, local or school level as data files, computer-generated listings, or by hand, depending on who has the most accurate information. Since it is important that the student sample be selected from an accurate complete list, the list needs to be prepared slightly in advance of the testing date. It is suggested that the list be received one to two months before the testing date so that the National Service Provider has adequate time to select the student sample.

*i. Preparing a list of age-eligible students*

A list of age-eligible students is to be prepared using the student tracking form that the OECD will provide. The following is considered important:

- The list is to include students who might not be tested due to a disability or limited language proficiency. Students who cannot be tested are to be excluded from the assessment after the student sample is selected. In calculating the response rate, if no replacement for the student is forthcoming, the number of missing students is excluded from the denominator.
- It is suggested that schools retain a copy of the student list in case the National Service Provider has to contact the school with questions.
- The student list is to be up-to-date at the time of the sampling rather than a list prepared at the beginning of the school year.
- Students are identified by their unique student identification numbers.
- Student grade and gender are to be reported on the student list. Date of birth is also captured, to check that a student will be within the eligible age range at the time of testing.

*ii. Student sample size*

For each school, a sample target size is set. This value is typically 85 students, although upon prior agreement, schools can use alternative values. When the school has more than 85 eligible students, a random selection of 85 students should be made. However, when the school has fewer than 85 students (but no less than 35), no random sampling is needed and all students should be included for testing. The minimum number of participants is 35. Schools that have a total of between 35 and 85 eligible students must test all students. Any exception to this standard requires prior approval from the OECD. To ensure that the minimum number of 35 students per school is reached, the National Service Provider is encouraged to test at least 45 students, assuming a participation rate of 80%.

*iii. Stratification*

Prior to sampling, students are to be stratified. Stratification consists of classifying students listed into strata (or groups) according to two selected variables referred to as stratification variables: student's gender and student's grade. Stratification is used in PFS to improve the efficiency of the sample design, thereby making the survey estimates more reliable, and ensure all types of students are included and adequately represented in the sample. The National Service Provider needs to compute the proportions of eligible students in each stratum (defined by the crossing of gender and grade), determines the target number of students to be sampled within each group given the total sample size targeted (85 students in general) and select random sample of the corresponding number



of students within each group. The recommended method for carrying out the selection is to use a ratio approach based on the expected total number of sampled students.

*iv. Preparing instruction for excluding students*

The PFS is a timed assessment administered in the instructional language of each participating school and is designed to be as inclusive as possible. However, students with limited assessment language experience or with physical, mental, or emotional disabilities should not be assessed. The National Service Provider uses the guidelines described in this report to develop any additional instructions. School Coordinator and Test Administrators need precise instruction for exclusions. The national operational definitions for within-school exclusion are to be clearly documented in the School Coordinator and Test Administrator manuals and submitted to the OECD for review before testing.

*v. Student replacements*

If it is likely that the response rate is going to be below 80%, due to parental refusals or absent students that are unlikely to attend a follow-up session, replacement students can be used. The replacement students are identified as follows. For each sampled student, the student immediately preceding and following it in the stratum, which was ordered within in the stratification process, are designated as his/her replacement students. The student immediately following the sampled student is designated as the first replacement, while the student immediately preceding the sampled student is designated as the second replacement.

*vi. Sending the student tracking form to the School Coordinator and the Test Administrator*

The School Coordinators need to know which students are sampled in order to notify students, parents and teachers, to update information, and to identify students to be excluded. The student tracking form should therefore be sent four weeks before the testing period. It is recommended that copies of the tracking form be kept by the National Service Provider and the Test Administrator in case the school copy is misplaced before the assessment day. The Test Administrator and School Coordinator manuals both assume that each would have a copy. In the interest of ensuring the PFS is as inclusive as possible, student participation and reasons for exclusion and replacement are separately coded in the student tracking form.

*vii. Student participation*

In order for a school to receive a report, a minimum of 35 examinees and a participation rate of 80% of the selected students are required. A student who participates in the original or make-up sessions is considered to be a participant. Students who do not reply to any of the cognitive items (i.e., all of whose responses to the cognitive items are

missing values) will be dropped from the scoring. Students can be considered participants as long as they reply to at least one cognitive item and there is enough information in the students' questionnaires to compute the economic, social and cultural status (ESCS) of at least 80% of the students tested in each school. Otherwise, such students are considered non-participating.

These requirements have been set to ensure adequate accuracy in estimating the school's results. Schools that do not meet these requirements will not be allowed to receive a report. National Service Providers must be very explicit in their communications to the schools about the requirements regarding student participation. Informing schools about this requirement before testing is strongly recommended.

### ***c. Packaging and shipping materials***

Regardless of how materials are packaged and shipped, the following need to be sent either to the Test Administrator or to the school:

- test booklets and student questionnaires for the number of students sampled;
- a student tracking form;
- a session attendance form; and
- additional student questionnaires and a bundle of extra test booklets.

It is also recommended that the National Service Provider send material reception and return forms to keep tallies of the materials.

Of the seven separate test booklets, one is pre-allocated to each student from a random starting point in each school. It is recommended that labels be printed, each with a student identification number and test booklet number allocated to that identification, as well as the student's name if this is an acceptable procedure within the country. Two or three copies of each student's label can be printed, and used to identify the test booklet, the questionnaire, and a packing envelope if used. National Service Providers are allowed some flexibility in how the materials are packaged and distributed, depending on local circumstances. It is specified however that the test booklets for a school be packaged so that they remain secure.

### ***d. Receipt of materials by the National Service Provider after testing***

It is recommended that the National Service Provider establish a database of schools before testing begins in order to record the shipment of materials to and from schools, keep tallies of materials sent and returned, and monitor the progress of the materials throughout the various steps in processing booklets after the testing. It is also recommended that upon receipt of materials back from schools, the counts of completed and unused booklets be checked against the participation status information recorded on the student tracking form by the Test Administrator.

### *e. Coding of the cognitive test and of the context questionnaires*

This section describes PFS coding procedures. Overall, a substantial share of the cognitive items across reading, mathematics, and science domains requires manual coding by trained coders. It is crucial for comparability of results in a study such as the PFS that students' responses are scored uniformly from coder to coder, and from country to country. Comprehensive criteria for coding, including many examples of acceptable and unacceptable responses, prepared by the OECD, will be provided to National Service Providers in coding guides for each of the three domains: reading, mathematics, and science.

In setting up the coding of students' responses to open-ended items, National Service Providers have to carry out or oversee several steps:

- Adapt or translate the coding guides as needed and submit these to the OECD for verification.
- Recruit and train coders.
- Locate suitable local examples of responses to use in training and practice.
- Organise booklets as they are returned from schools.
- Code the students' responses to the test booklets.
- Code the students' responses to the Student Questionnaire.
- Code the school's responses to the School Questionnaire.

### *f. Coding the test booklets*

The coding of the PFS test booklets is to be carried out following a single coding design. It is recommended to organise coding so that all appearances of each cluster type involved in the coding are coded together. This arrangement entails coders working with several booklet types at the same time and requires space for partly coded booklets to be stored while other booklets are being worked on. However, organising the coding this way has the substantial benefits of more accurate and consistent coding (because training and coding are more closely linked), minimising effects of coder leniency or harshness (coders code across the range of participating students and schools).

### *i. Staffing*

National Service Providers are responsible for recruiting appropriately qualified people to carry out the coding of the test booklets and the context questionnaires. Pools of experienced coders from other projects can be called upon. It is not necessary for coders to have high-level academic qualifications, but they need to have a good understanding of either mid-secondary level mathematics and science or the language of the test, and to be familiar with ways in which secondary-level students express themselves. Teachers on

leave, recently retired teachers and senior teacher trainees are all considered to be potentially suitable coders.

People are also needed to unpack, check and assemble booklets into labelled bundles so that coders can respect the design for randomly allocating sets of booklets to coders.

*ii. Confidentiality form*

Before seeing or receiving any copies of PFS test materials, prospective coders are required to sign a confidentiality form, obligating them not to disclose the content of the PFS tests beyond the groups of coders and trainers with whom they will be working.

*iii. Training*

Coders are required to attend one coder training session coordinated by the National Service Provider. At the training session, National Service Providers familiarise coding staff with the coding guides and their interpretation. It is recommended that prospective coders be informed at the beginning of training that they will be expected to apply the coding guides with a high level of consistency and that reliability checks will be made frequently by the overall supervisor as part of the coding process.

*iv. Length of coding sessions*

Coding responses to open-ended items is mentally demanding, requiring a level of concentration that cannot be maintained for long periods of time. It is therefore recommended that coders work for no more than six hours per day on actual coding, and take two or three breaks for coffee and lunch.

***g. Student questionnaire coding***

Questions from several PISA Student Questionnaires were assembled to form the PFS Student Questionnaire. The PFS Student Questionnaire includes most of the core items from the PISA 2009 Student Questionnaire to allow for direct comparisons with PISA results and analyses from 2009 onwards. Most of these core questions have remained unchanged in PISA 2015 and are part of a pool of basic questions that are retained for all PISA cycles.

The PISA Student Questionnaire Codebooks must be used to code student response to the PFS Student Questionnaire. The question IDs clearly indicate the original questionnaire the question stems from (see Section 3.a in Part 1). For instance, the first question: “What <grade> are you in?” is indexed as “15-ST001” meaning that it corresponds to question 1 of the PISA 2015 questionnaire. Therefore, responses to this question can be coded following the PISA 2015 Codebook.

The questions belonging to the pool of common PISA questionnaire questions must be coded using the most recent PFS Student Questionnaire codebook.

### ***h. School questionnaire coding***

The PFS School Questionnaire both in the reduced and extended versions includes questions common to the PISA 2009 and 2012 School Questionnaires. The PISA 2012 School Questionnaire Codebook must be used to code schools' responses to the reduced version of the PFS School Questionnaires. The PISA 2012 School Questionnaire Codebook must be privileged to code schools' responses to the extended version of the PFS School Questionnaires. Questions specific to the PISA 2009 School Questionnaire must be coded using the PISA 2009 School Questionnaire Codebook.



## Chapter 3

### Data Processing

For the data processing, the National Service Provider co-ordinates activities with analysts proficient in data processing and statistics, referred to as Statisticians. Statisticians are responsible for constructing the school datasets, the context and performance variables, and for generating schools' results delivered in the school report. Creating context and performance variables requires using scaling software. This report presents syntax examples of the scaling software ConQuest® developed by the Australian Council of Educational Research (ACER). Statisticians can work with another scaling software, such as the TAM package under R (a free open-source statistical software), after approval of the OECD. Results reported in the school report must be generated using standard methodologies and procedures. This part of the report also presents and refers to syntax and macros for statistical software packages such as SAS® and SPSS®, initially developed for analysing PISA data. SAS® and SPSS® syntax and macros are detailed in the second edition of the *PISA Data Analysis Manual* (OECD, 2009a and 2009b) for SAS® and SPSS® users. Statisticians are encouraged to ask the OECD any query or question regarding the implementation of the statistical procedures.

The National Service Provider will prepare the initial PFS databases for students and schools. The data file templates and codebooks will be provided by the OECD. The National Service Provider will prepare the data file according to the data file templates and codebooks, including a School Unique Identifier (14-digit code) created by concatenating the Country Code (3-digit code; for example, 101 USA, 102 UK, etc.), the School Identifier (9-digit code maximum, for example the School ID and the District number) and the Year of Administration (2-digit code, for example, 16 for 2016 and 17 for 2017). This School Unique Identifier must be exclusive for each school and for each participation year, making it possible to track schools across cycles. Special attention must be paid to schools that participate more than once, in which case the code should be the same except for the last two digits, which represent the Year of Administration.

#### 7. Constructing the initial PFS database

##### a. Files in the initial PFS database

The initial PFS database consists of three data files: two with student responses (raw data and scored data) and one with school responses, with all the information necessary to create the school report.

*i. Student data file*

For each student who participates in the assessment, the student data file must contain the following information:

- Identification variables for the school (School Unique Identifier) and the student.
- Test booklet identification (and possibly test language).
- The student responses to the cognitive items. Attention must be paid to the format of the responses for use in scaling software. For instance, when the scored responses consist of multiple digits (complex multiple choice or open-ended items), the multiple digits must be recoded into single-digit variables for use in the scaling software ConQuest®.
- The student responses to the Student Questionnaire.

The PFS items are organised into units. Each unit consists of a stimulus (consisting of a piece of text or related texts, pictures or graphs) followed by one or more questions. A unit is identified by a short label. The units' labels consist of six characters and form the first part of the item names in the data files. The first two characters are PR, PM or PS for reading, mathematics or science, respectively. The next four characters indicate the unit within the domain. For example, PM5124 is a mathematics unit. The item names (usually eight or nine digits) represent questions within a unit and are used as item names (in the current example the item names within the unit are PM5124Q01, PM5124Q02, PM5124Q03). Thus, items within a unit have the same initial six characters plus a question number. Responses that need to be recoded into single-digit variables have a "T" or "D" at the end of the variable name (e.g., PS7012Q07T).

For use in the scaling software ConQuest®, the scored data file typically includes one single digit variable per item with scores instead of response categories. In both the original and scored responses files, it is recommended that the cognitive items are sorted by domain and alphabetically and numerically by item name within domain. This means that the mathematics items appear at the beginning of the file, followed by the reading items and then the science items. Within domains, units with smaller numeric identification appear before those with larger identification, and within each unit, the first question will precede the second, and so on.

The National Service Provider will prepare and send to the OECD two versions of the student data file for revision: i) a raw student data file with the direct responses of the students to all the items (for example, 0, 1, 2 for cognitive items and 1, 2, 3, 4 for questionnaire items), ii) a scored student data file with all the final scores in Math, Reading and Science, the ESCS and the rest of the indicators from the questionnaire, such as self-efficacy and instrumental motivation.



### *ii. School file*

The school questionnaire data file must contain the following information for each school that participated in the assessment:

- Identification variables for the school (School Unique Identifier) and the student.
- The school responses on the (reduced or extended version of the) School Questionnaire.
- The testing date.

The school file must also contain the original variables collected through information provided by the school prior to testing, such as the number of students eligible for participating in the PFS. The school file must contain all information used in generating the school report. The National Service Provider will send these data files (that is, two versions of the student data file and the school data file) to the OECD for revision at least two weeks before the school report is expected to be delivered.

## ***b. Records in the database***

### *i. Records included in the database*

The student scored response file must contain records of all PFS students who attended test (assessment) sessions and replied to at least one cognitive item. PFS students who only attended the questionnaire session are included in the student questionnaire file if they provided at least one response to the student questionnaire, their sex is known, and the father's or the mother's occupation is known from the student questionnaire. The school file should contain records of all participating schools with 35 or more students participating in the assessment sessions.

### *ii. Records excluded from the database*

Students who do not reply to any of the cognitive items (i.e. all of whose responses to the cognitive items are missing) will be dropped from the sample and the scoring process. Students can remain in the database as long as they have replied to at least one cognitive item. Any other scenario in which the final number of valid students is lower than the minimum participation required will be studied case by case, based on evidence that removing these students does not introduce bias in the assessment (for example, removing low performers who prefer not to answer rather than to answer wrongly).

The following records should also be excluded from the student file:

- Sampled students who are reported as not eligible, students who are no longer at school, students who are excluded for physical, mental or linguistic reasons, and students who are absent on the testing day.

- Students who refused to participate in the assessment sessions.

Schools where fewer than 35 students participate in the assessment sessions or where the participation rate is lower than 80% should have their records excluded from the school file.

### Box 1. Examples of participating schools eligible to receive a school report

Term	Definition	Example 1: Standard school	Example 2: Participation issue	Example 3: Missing responses
Total number of eligible students	Number of students in the school who are eligible for testing	115	115	115
Sampled	Number of students in the school who are sampled to test	85	85	85
Participated <sup>1</sup>	Number of sampled students who show up to test	75	35	75
Valid <sup>2</sup>	Number of participating students who provide enough data to produce results	65	30	50
Criteria for delivering a report		Yes	No	No
Reason			Participation rate lower than 80% of students sampled	Number of valid responses lower than 80% of students sampled

Notes:

<sup>1</sup>In order to ensure that the minimum number of 35 students per school is reached, since frequently not all students sampled show up to test, the National Service Provider is encouraged to test at least 45 students, assuming a participation rate of 80%.

<sup>2</sup>The number of valid students includes those who have replied to at least one cognitive item and where student questionnaires include enough information to compute the economic and social cultural status (ESCS) of at least 80% of the students tested in each school. The number of valid students should be at least 80% of the students sampled.

Any exception to these standards will need prior approval from the OECD.

### c. Representing missing data

The coding of the data must distinguish between four different types of missing data:

- Item level non-response: 9 for a one-digit variable, 99 for a two-digit variable, 999 for a three-digit variable, and so on. Missing codes are shown in the codebooks. This missing code is used if the student or school principal is expected to answer a question, but no response is actually provided.

- Multiple or invalid responses: 8 for a one-digit variable, 98 for a two-digit variable, 998 for a three-digit variable, and so on. For the multiple-choice items code, 8 is used when the student selects more than one alternative answer.
- Not-administered: 7 for a one-digit variable, 97 for a two-digit variables, 997 for a three-digit variable, and so on. Generally, this code is used for cognitive and questionnaire items that are not administered to the students and for items that are deleted after assessment because of misprints or translation errors.
- Not reached items: all consecutive missing values clustered at the end of test session are replaced by the non-reached code, 'r', except for the first value of the missing series, which is coded as item level non-response.

#### *d. Merging the data files*

Once the data files are well prepared and organised, the student scored responses file must be merged with the student and school questionnaires data files using the student and School Unique Identifier variables. The following procedures are to be performed in data cleaning:

- Resolving cases of unmatched students or schools in the three data files.
- Removing unresolved cases of unmatched students or schools in the initial PFS database.

## **8. Procedures for scaling cognitive data**

Procedures for scaling cognitive data must be used in two cases: When assessing the psychometric properties of items administered during the field trial and when constructing student performance variables for reporting purposes. In both cases, Statisticians must use a specific form of a generalised item response model; the mixed coefficients multinomial logit model. This section first provides, in technical terms, an overview of the methodological background of this model. It describes the specific model used for analysing items from the field trial and the one used for computing students' performance scores at PFS. It then details the practical procedures for performing the estimation of such models, the item analyses required on field trial data and the construction of students' performance scores and proficiency levels.

#### *a. The mixed coefficients multinomial logit model*

As for main PISA surveys, the mixed coefficients multinomial logit model as described by Adams, Wilson and Wang (1997) must be used to scale the PFS cognitive data. This model results from the combination of an item response model and a population model. This section presents in rather technical terms the features of the general model and describes the two specific forms used for item analyses and for the construction of performance variables.

*i. The item response model*

As the PFS tests contain both dichotomous items (having two possible scores) and polytomous items (having more than two possible scores), the item response model used is a generalised form of the Rasch model. In that model, the probability that a student with ability  $\theta$  will obtain a score of  $j$  on item  $i$  is expressed as

$$P_{ij}(\theta) = P(X_{ij} = 1 | \theta, \Delta) = \frac{\exp(\theta - \delta_i + \tau_{ij})}{\sum_{k=0}^{M_i} \exp(\theta - \delta_i + \tau_{ik})}, j = 0, 1, \dots, M_i$$

Item  $i$  has  $M_i$  “steps”, and  $j$  is the number of “steps” successfully completed by the students or the number of credits obtained. The  $M_i + 1$  values range from 0 (not steps completed or no credit) to  $M_i$  (all steps completed or full credit obtained), and the steps are ordered; that is, for a given item, a higher score (a higher value of  $j$ ) reflects higher ability.  $\theta$  denotes the person’s latent trait, the item parameter  $\delta_i$  gives the location of the item on the latent continuum or difficulty of the item and  $\tau_{ij}$  are called step difficulties:  $\delta_{ij}$  is the difficulty of step  $j$  of item  $i$ . The overall scale is anchored using  $\sum_{i=1}^n \sum_{k=0}^{M_i} \delta_{ik} - \tau_{ik} \equiv 0$ , where  $n$  is the total number of items used for scaling.

The probability of obtaining a particular score on a particular item can be generalised to the probability of a response pattern to all items of the domain. For each item  $(i)_{i=1, \dots, n}$ , The  $M_i + 1$  variables  $(X_{ij})_{j=0, 1, \dots, M_i}$  are collected together into a vector  $X_i$ . The  $n$  vectors  $X_i$  are collected together into a single vector  $X$ , called the response pattern. The probability that a student with ability  $\theta$  will obtain a particular response pattern  $\{x\}$  is thus modelled as

$$f(x | \theta, \Delta) = P(\{X = x\} | \theta, \Delta) = \prod_{i=1}^n \prod_{j=1}^{M_i} P_{ij}(\theta)^{x_{ij}}$$

which can be rewritten in a simplified form as

$$f_x(x | \theta) = \frac{\exp(x(\theta - \Delta))}{\sum_{u \in \Omega} \exp(u(\theta - \Delta))}$$

where  $\Omega$  is the set of all possible response patterns and  $\Delta$  is a vector containing sums of item parameters.

When using the item response model for analysing item characteristics, Statisticians must estimate item difficulty parameters ( $\Delta$ ) from item response data of the field trial. Those parameters are the outcomes of interest for the item analyses. When using the item response model for estimating students’ ability ( $\theta$ ), Statisticians must anchor (or fix) item difficulty parameters at their international values (provided by the OECD). In this case, students’ ability estimates are the outcomes of interest.

### ii. The population model

The item response model is a conditional model, in the sense that it describes the process of generating item responses conditional on the latent variable,  $\theta$ . The complete definition of the model, therefore, requires the specification of a density,  $f_{\theta}(\theta)$ , for the latent variable,  $\theta$ . Student latent ability is assumed to be sampled from a normal population with mean  $\mu$  and variance  $\sigma^2$ . That is:

$$f_{\theta}(\theta; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta - \mu)^2}{2\sigma^2}\right)$$

Or equivalently  $\theta = \mu + \varepsilon$  where  $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ .

This is how the population model must be specified when analysing item response data from the field trial.

Adams, Wilson and Wang (1997) discuss how a natural extension of this equation is to replace the mean,  $\mu$ , with the regression model,  $Z\beta$ , where  $Z$  is a vector of fixed and known values for a student, and  $\beta$  is the corresponding vector of regression coefficients. The population model can be re-written as:

$$f_{\theta}(\theta; \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta - Z\beta)^2}{2\sigma^2}\right)$$

The  $Z$  variables are referred to as conditioning variables. If this equation is used as the population model then the parameters to be estimated are  $\beta$  and  $\sigma^2$ .

This is how the population model must be specified when estimating students' ability.

### iii. The combined model

The conditional item response model and the population model are then combined to obtain the unconditional, or marginal, item response model.

The combined model used for the item analyses of the field trial data can be written as:

$$f_X(X; \Delta) = \int_{\theta} f_X(X; \Delta | \theta) f_{\theta}(\theta; \beta, \sigma^2) d\theta \quad (\text{Equation 1})$$

This combined model was initially used for estimating the international item parameters ( $\Delta$ ) during the international pilot and equating study. Values of the international item parameters will be provided by the OECD.

The combined model used for estimating students' ability can be written as:

$$f_X(X | \Delta) = \int_{\theta} f_X(X | \theta, \Delta) f_{\theta}(\theta; \beta, \sigma^2) d\theta \quad (\text{Equation 2})$$

It is important to recognise that under this model the locations of individuals on the latent variable ( $\theta$ ) are not estimated. For each individual it is possible, however, to specify a posterior distribution for the latent variable, given by:

$$h_{\theta}(\theta|X, \Delta) = \frac{f_X(X|\theta, \Delta)f_{\theta}(\theta; \beta, \sigma^2)}{f_X(X|\Delta)}$$

When using this combined model to estimate students' abilities, item parameters  $\Delta$  must be anchored at their international values. The posterior distribution of the latent student ability must be used to generate plausible values of each student's score.

*iv. Plausible values*

As with all item response scaling models, student proficiencies are not observed; they are missing data that must be inferred from the observed item responses. There are several alternative approaches for making this inference. As in main PISA, PFS uses the imputation methodology usually referred to as plausible values (PVs). PVs are a selection of likely proficiencies for students that attained each score. For each scale, five plausible values per student are included in the database.

Using item parameters anchored at their estimated values from the international calibration (provided by the OECD), the plausible values are random draws from the marginal posterior of the latent distribution for each student. For details on the generation of plausible values, see Chapter 9 of the *PISA 2012 Technical Report* (OECD, 2014). Drawing plausible values in PFS data analysis can easily be implemented using statistical software suited for item response model estimation.

***b. Application to PFS***

National service providers (NSP) must use two different combined models: One for performing national item analyses on field trial data, the other for generating students' scores. In both cases, Statisticians must use scaling software which can perform generalised item response models.

*i. Item analyses of the field trial data*

National item analyses of the field trial data must be performed separately, country by country, and within country, language by language, using unweighted data. For the item analyses, the conditional item response model must be used in conjunction with the population model without conditioning variables

$f_X(X; \Delta) = \int_{\theta} f_X(X; \Delta|\theta)f_{\theta}(\theta; \beta, \sigma^2)d\theta$  (Equation 1). That is, it will be assumed that students have been sampled from a normal distribution. In the item response part of the model, the item parameters will be freely estimated and not anchored at their international values.

An example of a ConQuest® syntax used to fit partial credit items and run the partial credit sample analysis on students' responses to the reading test is presented in Box 2. The datafile PFS\_scored.dat read by ConQuest® must contain the scored responses and not the original responses of the students. The scored responses to the 46 reading items

are placed in the ascending order of the item identifiers in column 19 through 64 of the datafile. The *codes* statement restricts the list of codes that ConQuest will consider valid to 0, 1 and 2. Any other codes for the items will be treated as missing-response data. The *model* statement contains two terms (*item* and *item\*step*) and results in the estimation of two sets of parameters. The term *item* results in the estimation of a set of item difficulty parameters, and the term *item\*step* in a set of item step-parameters that vary across items. This is the generalised partial credit model. The *estimate* statement is used to initiate the estimation of the model. The *export* statement produces a display of the item response model parameter estimates and saves them to the file *readnat.prm*. The option *estimates=latent* requests that the displays include an illustration of the latent ability distribution and redirect them to the file *readnat.shw*. The *itanal* statement produces a display of the results of a traditional item analysis and redirects them to the file *read.itn*.

### Box 2. ConQuest® syntax for running the item analyses on the reading test

```
Title Item analyses for reading;
datafile PFS_scored.dat;
format id 1-5 responses 19-64;
labels << read.lab ;
codes 0,1,2;
model item + item*step;
estimate;
export parameters >> readnat.prm;
show ! estimates=latent >> readnat.shw;
itanal >> read.itn ;
quit;
```

The outcomes of the item analyses are used to make a decision about how to treat each item in the participating country and for each language. This means that an item may be deleted from the scaling in a particular country and in particular languages if it has poor psychometric characteristics in this particular country and those particular languages. When reviewing the national item analyses, particular attention must be paid to the fit of the items to the scaling model, item discrimination and item-by-country or language interactions. Four types of item analyses must be carried out by the Statistician(s): item response model fit, discrimination coefficients difference in national and international item difficulties and differential item functioning (DIF) by gender. The outcomes of these four analyses, a national list of ‘dodgy’ items and recommendations regarding item treatment must be reported to the OECD for approval.

### Item response model fit (weighted mean square MNSQ)

For each item parameter, the fit mean square index (Adams, Wilson and Wang, 1997) must be used to provide an indication of the compatibility of the model and the data. For each student, the model describes the probability of obtaining the different item scores. It is therefore possible to compare the model prediction and what has been observed for one item across students. Accumulating comparisons across students gives an item-fit statistic. As the fit statistics compare an observed value with a predicted value, the fit is an analysis of residuals. In the case of the item infit mean square, values near one are desirable. A weighted MNSQ greater than one is associated with a low discrimination index, meaning the data exhibits more variability than expected by the model, and an infit mean square less than one is associated with a high discrimination index, meaning the data exhibits less variability than expected by the model. As in main PISA, PFS accepts small variation of MNSQ around one. Items with values of weighted MNSQ larger than 1.2 and below 0.8 must be reported as dodgy items.

### Discrimination coefficients

For each item, the correlation between the student's score on the item and the student's aggregate score on the set for the same domain as the item of interest must be used as an index of discrimination. If  $p_{ij}$  (calculated as  $x_{ij}/m_i$ ) is the proportion of score levels that student  $i$  achieved on item  $j$ , and  $p_i = \sum_j p_{ij}$  (where the summation is of the items from the same domain as item  $j$ ) is the sum of the proportions of the maximum score achieved by student  $i$ , then the discrimination is calculated as the product-moment correlation between  $p_{ij}$  and  $p_i$  for all students. Items with a discrimination coefficient below 0.2 must be reported as dodgy items.

### Difference in item difficulties

The national scaling provides nationally specific item parameter estimates. The consistency of item parameter estimates across countries is of particular interest for PFS. If the test measures the same latent trait per domain in all participating countries, then nationally specific items should have the same relative difficulty or, more precisely, would fall within a specific confidence interval.

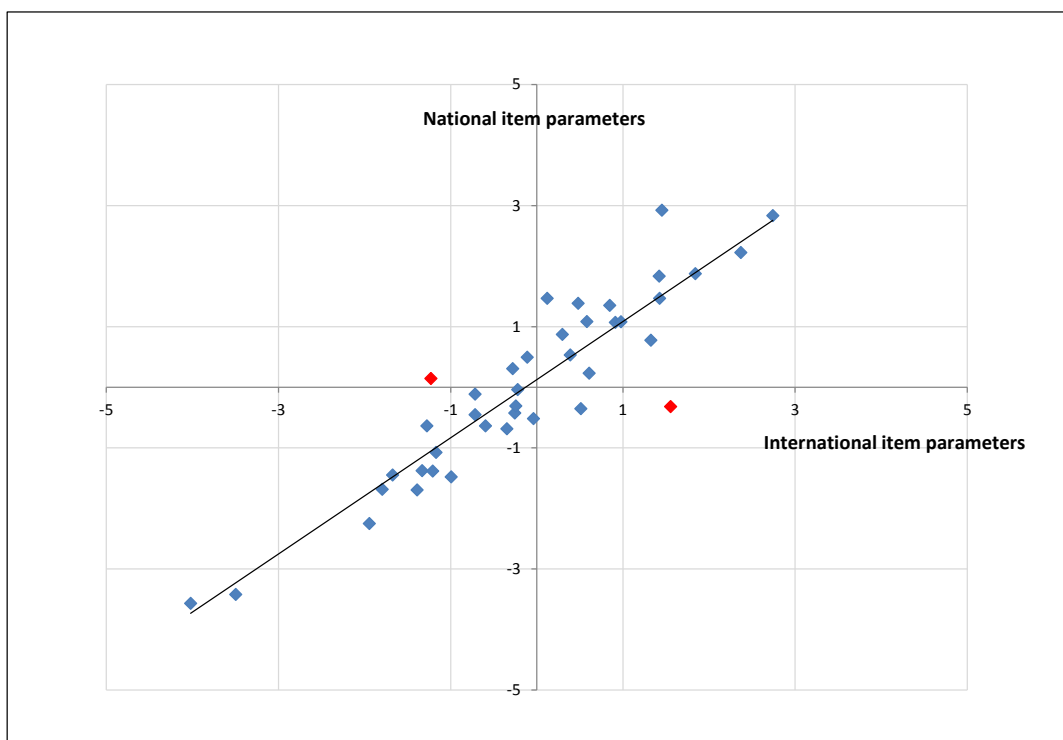
Statisticians are asked to report scatter plots of national and international PFS item difficulties per domain. International PFS item difficulties will be provided by the OECD. Both sets of difficulties are centred on zero and are therefore referred to as relative difficulties. Both sets of difficulties must also be standardised with a standard deviation of one. The vertical axis must represent the national standardised item difficulties and the horizontal axis the OECD or international standardised item difficulties. Each item must be represented by a dot. A linear regression line must be added to the scatter plot to illustrate the relationship between notational and international item difficulties.

Figure 1 provides an overview of the behaviour of all items in a domain in one country compared to the pooled international set and assists in identifying outliers. Items



that lie exactly on the regression line have similar national and international relative item difficulties. An outlier occurs when the relative national item difficulty is very different from the predicted one given the OECD/international relative item difficulty. For each item, Statisticians will calculate the difference between the international and national difficulties. Each difference will standardise the difference by dividing each of them by the standard deviation of all differences. Statisticians will then list all items with a standardised difference of more than 1.96. In Figure 1 there are a couple of obvious outliers (red dots). This suggests that something could be wrong with these items. They must be reported as dodgy in the report.

**Figure 1. Example of scatter plot**



### Differential Item Functioning

Differential item functioning (DIF) analysis is central to the process of psychometric validation of tests and questionnaires. International guidelines on educational measurement and test development demand that DIF analysis be carried out to ensure construct equivalence (International Test Commission, 2013). Statisticians are asked to perform DIF analysis in order to detect any potential bias between the girls and boys responding. Within the context of Rasch modelling, an item is deemed to exhibit DIF if the response probabilities for that item cannot be fully explained by the ability of the student and a fixed set of difficulty parameters for that item.

The DIF analysis is performed using the multi-facet model of ConQuest (Wu et al., 2007) through the difference in parameters of item characteristic curves (ICCs) of groups of girls and boys. The DIF value for each item is computed as the difference between the

two relative difficulty estimates (boys versus girls). The corresponding chi-square test for this difference is obtained from the DIF value and the standard errors of the estimates. An item is flagged as having substantial DIF if the chi-square DIF test is significant at a 0.01 level and the absolute DIF value is greater than 0.20 logits. A negative gender DIF value means that the item relatively favours girls, and if the gender DIF is positive, the item relatively favours boys.

### National list of dodgy items

For each language, National Service Providers must report all the items identified as dodgy and the criteria these items do not satisfy. They are asked to check them carefully for any translation or printing errors. After the verification of the dodgy items, National Service Provider must recommend decisions regarding their treatment for the student score generation for the field trial data and for future administration of the survey. It is recommended that items with translation and misprint problems be discarded for the student score generation from the field trial data and improved for future administration of the survey. Once the report containing the national item analyses has been reviewed and decisions regarding dodgy items approved by the OECD, Statisticians can proceed with the student score generation.

### *ii. Student score generation*

A school report presents a school's results from the PFS and compares the students' performances in three subjects (reading, mathematics and science) with performances of peers in countries and economies that took part in past PISA surveys. For reporting purposes, thirty performance variables must be constructed and included in the database: five plausible values of student's performance score and five plausible levels of student's proficiency for each of the three domains of assessment.

### Omitted responses

The PISA-based Test for Schools will treat the number of missing responses with a differentiation between item level non-responses and not-reached responses. A response is coded as an item level non-response if the student was expected to answer a question but provided no response. All consecutive missing values clustered at the end of a test session are replaced by the not-reached code, except for the first value of the missing series, which is coded as an item level non-response. Therefore, for the single missing value at the end of the test session and for the first missing value of consecutive missing values (i.e.,  $\geq 2$  missing values) at the end of the test session, the missing values are coded as missing. All non-first missing values in a missing series at the end of the test session are coded as not-reached.

The number of not-reached items is used in PISA 2015 as a source of background information in the generation of plausible values, so that the correlation of not-reached items and proficiency is modelled and accounted for in the results. This change makes the treatment of student responses consistent across the estimation of item parameters and

student proficiency, and eliminates potential advantages for test takers who randomly guess answers to multiple-choice questions that they could not complete in time, as contrasted with test takers who leave these non-reached items unanswered. This new treatment of non-reached items might result in higher scores than would have been estimated in the past for students who leave many items unanswered. The PISA-based Test for Schools will accommodate this change as follows: the ratio of not-reached items (number of non-reached/total of items in the booklet) will be introduced as a direct regressor in the estimation of the plausible values, in addition to the rest of the student-level information. A further explanation of the process can be found in the following paragraphs.

### Estimation of the plausible values

Estimation of the combined model presented in Equation 2 must be performed separately for each domain of assessment. Fifteen plausible values, five for each of the three PFS scales, must be included in the PFS database: MATHPV1 to MATHPV5 are for mathematical literacy, READPV1 to READPV5 for reading literacy and SCIEPV1 to SCIEPV5 for scientific literacy.

The PFS conditioning variables (vector  $Z$  in Equation 2) are prepared using procedures based on those used in PISA. All available student-level information, other than their responses to the items in the booklets, is used either as direct or indirect regressors in the population model. The preparation of the variables for the conditioning proceeds as follows. Variables for booklet identifier are represented by deviation contrast codes and are used as direct regressors. Each booklet is represented by one variable, except for Reference Booklet 7. Booklet 7 was chosen as the reference booklet because it includes items from every domain. The difference between the simple contrast codes that were used in PISA 2000 and PISA 2003 was that with deviation contrast coding, the sum of each column is zero, whereas for simple contrast coding, the sum is one. The contrast coding scheme is given in Annex B of PISA's Technical Reports. Further information can be found in PISA 2012 Technical Report (p. 157). Using this method, the imputation of abilities for students who did not respond to any science or reading items is based on information from all booklets that have items in a domain, and not simply from the reference booklet, as in simple contrast coding. Other direct variables in the regression are gender (and missing gender, if any entries are missing), grade, parents' highest occupational status (HISEI) (see Section 9.b), school dummy variables (with the largest school as a reference; '-1' in all dummies), and the ratio of not-reached items.

All other categorical variables from the Student Questionnaire are dummy coded. Four numerical variables must be recoded as follows:

$$\begin{aligned} \text{AGE} &= (100 + \text{test year} - \text{student's birth year}) + (\text{test month} - \text{student's birth month})/12 \\ &= (\text{test year} - 15\text{-ST003b}) + (\text{test month} - 15\text{-ST003a})/12 \end{aligned}$$

All dummy variables, the numerical variables age of school entry (09-ST06), age of arrival in the country (15-ST21), as well as the recoded numerical variable (AGE) are analysed in a principal component analysis. The number of component vectors that must be extracted and used in the scaling model as regressors is country specific and must explain 95% of the total variance in all the original variables. Therefore, as many components as are needed to explain 95% of the variance must be extracted. This indicates that the number of components can be different in each country, depending on the influence of the regressors on the estimation of the plausible values. Whenever possible, standardised regressors should be introduced to make the interpretation of the outcomes easier.

The item-response model is fitted for each domain of assessment to each national PFS data set and the model parameters are estimated using item parameters anchored at their international location (provided by the OECD), and the direct conditioning variables as well as the extracted component vectors described above must be introduced in the regression. A minimum number of 2000 observations are required for the model to be correctly fitted. Five plausible values must be drawn from the multivariate normal distribution following the methodology described earlier in this section. At this stage, the plausible values drawn are expressed in logits and must still be put onto PISA point scales.

An example of the ConQuest® syntax used for estimating the plausible values in reading is presented in Box 3. The datafile PFS\_scored.dat read by ConQuest® must contain the scored responses and not the original responses of the students. The scored responses to the 46 reading items are placed in the ascending order of the item identifiers in column 19 through 64 of the data file. The reading test contains dichotomous items (scored 0 or 1) and polytomous items (scored 0, 1 or 2) so the *codes* statement limit the valid codes to 0, 1 and 2. The vectors extracted from the principal component analysis are named PCAvec1, PCAvec2, PCAvec3, and so on. School dummies are named school001, school002, school003, and so on. In this example, it is assumed that data from 40 schools are being scaled. For the sake of brevity, only four school dummies are listed in the *format* command. In the example, it is assumed that 3 principal component vectors have been extracted. The *regression* command must contain the list of all direct regressors and all vectors extracted from the principal component analysis. Booklets 4 and 7 are excluded from the direct regressors as booklet 4 contains no reading item and because booklet 7 is chosen as the reference booklet. The *model* statement contains two terms (*item* and *item\*step*) and results in the estimation of two sets of parameters. The term *item* results in the estimation of a set of item difficulty parameters, and the term *item\*step* in a set of item step-parameters that vary across items. The *estimate* statement is used to initiate the estimation of the model. The anchored values of the item parameters (the international item parameters) must be imported. The order of the anchored parameters in the read.prm file must fit with the order of the item responses. The read.prm file contains 49 values: the first 46 values correspond to the difficulty parameters of each of the items; the last 3 values are the step-parameters of the three polytomous reading items

(PR6014Q06T, PR6017Q05T and PR6018Q08). The file read.pls will finally be created when the command file is executed. It will contain 5 plausible values for each case.

### Box 3. ConQuest® syntax for generating the plausible values in reading

```
Title Plausible values reading;
datafile PFS_scored.dat;

format id 1-5 Schoolid 6-49 female 50 male 51 gender_m 52 grade 53 booklet1 54 booklet2
55 booklet3 56 booklet4 57 booklet5 58 booklet6 59 booklet7 60 HISEI 61-65 PCAvec1 66-75
PCAvec2 76-85 PCAvec3 86-95 school001 96 school002 97 school003 98 school040 125;

labels << read.lab ;

codes 0,1,2;

regression female male grade booklet1 booklet2 booklet3 booklet5 booklet6 HISEI not-
reachedratio PCAvec1 PCAvec2 PCAvec3 school002 school003 school040;

model item + item*step;

import anchor_parameters << read.prm;

export reg_coefficients >> read.reg;

export covariance >> read.cov;

set constraints=none,update=yes,warnings=no,n_plausible=5;

estimate ! fit=no;

show cases ! estimates=latent >> read.pls;

quit;
```

### Transforming the plausible values to PISA scales

For PISA surveys, the reading, mathematics and science results are each reported on the scales that were established when the respective domain was a major domain. For reading the reference scale was established for PISA 2000, for math it was for PISA 2003, for science it was for PISA 2006. So, for instance for PISA 2012, the reading results are thus reported on the PISA 2000 scale, the mathematics results on the PISA 2003 scale, the science results on the PISA 2006 scale. This was made possible because the new PISA tests were equated to the former PISA tests. In order to facilitate the interpretation of scores assigned to students, the PISA reading, mathematics and science reference scales were designed to have an average score of 500 points and a standard deviation of 100 across OECD countries.

For PFS the reading, mathematics and science results are also each reported on the reference PISA scales of the domain. This is also made possible because PFS tests were linked to the former PISA tests. Thanks to a transformation, students' scores at PFS can thus be expressed in score points and put on the reference PISA scale. The

transformations to put the PFS plausible values expressed in logits (PFS\_L) on the PISA point scales are:

- Reading:  $PISA2000 = 500 + 100 * (PBTS_L * .883 - 0.4837)/1.1002$
- Mathematics:  $PISA2003 = 500 + 100 * (PBTS_L + 0.1344)/1.2838$
- Science:  $PISA2006 = 500 + 100 * (PBTS_L - 0.1797)/1.0724$

### Generation of the proficiency levels

In order to render PISA results more accessible to educators, proficiency scales have been developed for the assessment domains. Since these scales are divided according to levels of difficulty and performance, both a ranking of student performance and a description of the skill associated with that proficiency level can be obtained. Each successive level is associated with tasks of increased difficulty. In PFS, six levels of proficiency, as defined in former PISA surveys, are used for each domain of assessment. The cut off points that frame the proficiency levels in reading, mathematics and science are presented in Table 4.

**Table 4. Lower score limits for the proficiency levels in reading, mathematics and science**

Domain	Lower score limit					
	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Reading	334.75	407.47	480.18	552.89	625.61	698.32
Mathematics	357.77	420.07	482.38	544.68	606.99	669.30
Science	334.94	409.54	484.14	558.73	633.33	707.93

Source: OECD (2014), *PISA 2012 Technical Report*, Table 15.1, p. 297; OECD (2012), *PISA 2009 Technical Report*, Table 15.1, p. 266; OECD (2009), *PISA 2006 Technical Report*, Table 15.1, p. 293.

Proficiency levels must be derived from the plausible values. Five plausible proficiency levels must be assigned to each student respectively according to their five plausible values. The SAS® and SPSS® syntaxes for the generation of the plausible proficiency levels are provided in chapter 9 of the *PISA Data Analysis Manual* (OECD, 2009a and 2009b).

## 9. Procedures for constructing contextual variables

### a. Overview

The PFS Student Questionnaire includes numerous items on student characteristics, student family background, and student perceptions. Responses to the items are transformed to be displayed in the school report. Some of the items are designed to be used in analyses as single items (for example, gender). However, most questionnaire items are designed to be combined in some way in order to measure latent constructs that

cannot be observed directly. To these items, transformations or scaling procedures must be applied to construct meaningful indices.

This section describes how indices, used in the analyses of the school report, must be constructed. Three different kinds of indices can be distinguished:

- Simple indices: these indices must be constructed through the arithmetical transformation or recoding of one or more items.
- Scale indices: these indices are constructed through the scaling of several items. Typically, scale scores for these indices are estimates of latent traits derived through Item Response Theory (IRT) scaling of dichotomous or polytomous items.
- Complex indices: these indices are derived from several other indices.

All indices have been used in main PISA surveys and must be constructed following the same methodology. It is recommended that the Statisticians responsible for the construction of contextual variables first test their procedures on PISA data (for instance, the PISA data of their country if the country has already participated in PISA), by re-constructing individual scores for each contextual variable. Once the procedures allow replicating the PISA indices, they can be used to generate PFS individual scores for each contextual index.

### ***b. Simple questionnaire variables***

#### *i. Grade*

The relative grade index (GRADE) is computed to capture between-school variation. It indicates whether students are in the school's modal grade  $f$  (value of 0) or whether they are below or above the modal grade ( $+x$  grades,  $-x$  grades). The information about the students' grade level was taken from the student questionnaire (ST001), whereas the modal grade was defined by the country and documented in the Student Tracking Form.

#### *ii. Student age*

The age of a student (AGE) is calculated as the difference between the year and month of the testing and the year and month of a student's birth. Data on students' age were obtained from both the questionnaire (ST003) and the student tracking forms. If the month of testing was not known for a particular student, the median month for that country was used in the calculation. The formula for computing AGE was

$$\text{AGE} = (100 + T_y - S_y) + (T_m - S_m)/12$$

where  $T_y$  and  $S_y$  are the year of the test and the year of the student's birth, respectively, in two-digit format (for example 06 or 92), and  $T_m$  and  $S_m$  are the month of the test and month of the student's birth, respectively. The result is rounded to two decimal places.

### iii. Grade repetition

The grade repetition variable (REPEAT) is computed by recoding variables ST127Q01TA, ST127Q02TA and ST127Q03TA. REPEAT took the value of 1 if the student had repeated a grade in at least one International Standard Classification of Education (ISCED) level, and the value of 0 if “No, never” was chosen at least once, given that none of the repeated grade categories were chosen. The index is assigned a missing value if the student did not tick any of the three categories in any levels.

### c. Questionnaire scale indices

Some of the scales that have been implemented in the PISA 2015 questionnaires, such as science self-efficacy and instrumental motivation in science, can be linked to the respective scales administered in PISA 2006, via a common calibration-linking procedure. For this purpose, international item and person parameters were obtained from a Generalised Partial Credit Model (see Text Box 7 for further explanation) in a single analysis based on data from all students in all countries from both cycles (2006 and 2015) using the mdltm software (von Davier, 2008). For each scale, only students with a minimum number of three valid responses were included. Students were weighted using the final student weight, and each country in each cycle contributed equally to the estimation. Additional analyses on the invariance of item parameters across countries, languages and cycles were conducted and unique parameters were assigned if necessary. Once this process was completed, weighted likelihood estimates (WLEs) for all examinees were obtained and the OECD mean and standard deviation of the newly constructed WLEs were matched to the OECD mean and standard deviation of the original 2006 WLEs, by applying the linear transformation equation of the form

$$WLE_{2015}^* = A * WLE_{2015} + B$$

in which

$$A = \frac{SD_{2006,original}}{SD_{2006,new}}$$

$$B = M_{2006,original} - A \times M_{2006,new}.$$

This procedure links the 2015 data for each of the trend scales to the respective scale established in PISA 2006. The correlations between the original and new WLEs for PISA 2006 indicated that all scales could be satisfactorily recovered. This is particularly encouraging, since the scaling model changed from the Partial Credit Model in previous cycles of PISA to the Generalised Partial Credit Model in 2015. Further description of these analyses can be found in Chapter 16 of the PISA 2015 Technical Report.

PISA 2006 parameters estimated with the Partial Credit Model are therefore used in PFS for science self-efficacy and instrumental motivation in science parameters. The parameters of the other questionnaire scale indices will also remain the same as in



previous PFS cycles, obtained from the original PISA cycle (i.e. PISA 2012 parameters for instrumental motivation in mathematics). The only student questionnaire indicator that has new parameters is the ESCS (see next section for further explanation). In sum, the scaling methodology in PFS remains the same as for trend comparisons in PISA, making the analysis consistent between different cycles and comparable with PISA 2015.

### *i. Scaling methodology*

Questionnaire scale indices are derived from student responses to several items, using a scaling methodology similar to that used for generating students' performance scores. In the scaling model,  $\theta$  no longer represents student's ability in a domain but a student's latent trait. The mixed coefficients multinomial logit model described in Equation 1 of Section 8.a.iii (without conditioning) must be fitted to student responses data. When using the unconditional item response model to estimate students' latent traits, item parameters  $\xi$  must be anchored at their international values. In the following subsections, international item parameters (obtained from PISA calibration samples using the ConQuest® software) are provided for each index.

For each student, it is then possible to specify a posterior distribution for the latent trait, given by:

$$h_{\theta}(\theta; \xi, \mu, \sigma^2 | X) = \frac{f_X(X; \xi | \theta) f_{\theta}(\theta | \mu, \sigma^2)}{f_X(X; \xi, \mu, \sigma^2)}$$

This posterior distribution of student latent trait must be used to generate Weighted Likelihood Estimates (WLEs) of each student's latent trait (and not plausible values as for students' performance scores).

WLEs must finally be transformed to an international metric with an OECD average of zero and an OECD standard deviation of one. The transformation must be achieved by applying the formula:

$$\theta' = \frac{\theta - \bar{\theta}_{\text{OECD}}}{\sigma(\theta_{\text{OECD}})}$$

where  $\theta'$  is the student's trait score in the international metric,  $\theta$  the original WLE in logits, and  $\bar{\theta}_{\text{OECD}}$  is the OECD mean of logit scores with equally weighted country samples.  $\sigma(\theta_{\text{OECD}})$  is the OECD standard deviation of the original WLEs. Means and standard deviations used for the transformation into the international metric are shown in Table 5.

**Table 5. OECD means and standard deviations of Weighted Likelihood Estimates (WLEs)**

Student-level indices	PISA	Mean	S.D.
DISCLIMA (Testing language)	2009	-0.79	2.27
DISCLIM (Mathematics)	2012	0.97	2.01
DISCLISCI	2015	0.53	1.3
HOMEPOS	2015	0.66	0.53
INSTMOT	2012	1.32	2.78
LMINS	2012	214.69	93.93
MATHEFF	2012	1.15	1.50
STUDREL	2012	1.32	2.01
WEALTH	2012	1.25	1.10
SCIEEFF	2006	0.45	1.31
INSTSCIE	2006	0.65	3.19

Source: OECD (2014), *PISA 2012 Technical Report*, Table 16.1, p. 313.

### ii. Mathematics self-efficacy

Eight items are used in the PFS as well as in PISA 2012 to measure mathematics self-efficacy (MATHEFF). The four response categories are “Very confident”, “Confident”, “Not very confident” and “Not at all confident”. All items must be reversed, so the higher difficulty corresponds to the higher level of confidence. Table 6 shows the item wording and the PISA 2012 international item parameters for this scale. For this index, item difficulties ranged from a comparatively easy one “Solving an equation like  $3x+5=17$ ” ( $\delta = -0.62$ ) to more difficult ones, such as “Finding the actual distance between two places on a map with a 1:10 000 scale” ( $\delta = 0.56$ ) and “Calculating the petrol consumption rate of a car” ( $\delta = 0.65$ ). The scaling procedure of this index thus takes into account the fact that students feel more confident in solving linear equations than they feel applying rates and proportions to real life situations.

**Table 6. Item parameters for mathematics self-efficacy (MATHEFF)**

Item	How confident do you feel about having to do the following mathematics tasks?		Parameter estimates		
			Delta	tau_1	tau_2
ST37Q01	a )	Using a <train timetable> to work out how long it would take to get from one place to another	-0.16537	-1.85539	-0.10055
ST37Q02	b )	Calculating how much cheaper a TV would be after a 30% discount	-0.31699	-1.69833	-0.05907
ST37Q03	c )	Calculating how many square metres of tiles you need to cover a floor	0.13366	-1.83729	0.14817
ST37Q04	d )	Understanding graphs presented in newspapers	-0.15505	-1.78957	-0.12065
ST37Q05	e )	Solving an equation like $3x+5=17$	-0.62211	-1.09181	-0.05693
ST37Q06	f )	Finding the actual distance between two places on a map with a 1:10 000 scale	0.55600	-1.89647	0.29021
ST37Q07	g )	Solving an equation like $2(x+3) = (x+3)(x-3)$	-0.07740	-1.38321	0.06433
ST37Q08	h )	Calculating the petrol consumption rate of a car	0.64726	-1.96788	0.13774

Source: OECD (2014), *PISA 2012 Technical Report*, Table 16.11, p. 322.

Box 4 shows an example of a ConQuest® command that can be used for generating the WLEs of the mathematics self-efficacy (MATHEFF) scale and easily adapted for generating other scale indices. The file *matheff.dat* contains student scored responses to the eight items located in columns 5 through 12. The four *key* statements are used to do the reverse scoring of the eight items. The *model* statement for these data contains two terms (*item* and *item\*step*) and will result in the use of two sets of parameters: a set of item difficulty parameters and a set of item step-parameters. The *import* command identifies the file *matheff.prm* containing the anchored values for the item parameters. Box 5 shows the order of the item parameters in the file *matheff.prm*. The command file listed in Box 4 also produces the file *matheff.wle*. This file contains WLEs for each of the students in the file *matheff.dat*. The format of these files is as follows. The file *matheff.wle* contains one line of data for each student in the sample who provides a valid response to at least one of the eight items that are analysed. The first block of columns contains an identification number for the case, which is the sequence number of the student (*id*) in the original data file. The next blocks of columns contain: the total score that the student attained; the maximum possible score that the student could have attained; the WLE of the student's latent ability; and an asymptotic standard error for that ability estimate. An extract from *matheff.wle* is shown in Box 6. Further detail about ConQuest® procedures can be found in the ConQuest Manual developed by ACER (Adams, Wu and Wilson, 2012).

**Box 4. ConQuest® syntax for generating the WLEs of the MATHEFF scale**

```
datafile matheff.dat ;
format id 1-4 responses 5-12 ;
key 11111111 ! 3 ;
key 22222222 ! 2 ;
key 33333333 ! 1 ;
key 44444444 ! 0 ;
model item + item*step ;
set constraints = none ;
import anchor_parameters << matheff.prm ;
estimate ;
show cases ! estimates = wle >> matheff.wle ;
show ! estimates = latent >> matheff.shw ;
itanal >> matheff.itn ;
```

**Box 5. Contents of the item parameter file matheff.prm**

1	-0.16537	/* ST37Q01 */
2	-0.31699	/* ST37Q02 */
3	0.13366	/* ST37Q03 */
4	-0.15505	/* ST37Q04 */
5	-0.62211	/* ST37Q05 */
6	0.55600	/* ST37Q06 */
7	-0.07740	/* ST37Q07 */
8	0.64726	/* ST37Q08 */
9	-1.85539	/* ST37Q01 step 1 */
10	-0.10055	/* ST37Q01 step 2 */
11	-1.69833	/* ST37Q02 step 1 */
12	-0.05907	/* ST37Q02 step 2 */
13	-1.83729	/* ST37Q03 step 1 */
14	0.14817	/* ST37Q03 step 2 */
15	-1.78957	/* ST37Q04 step 1 */
16	-0.12065	/* ST37Q04 step 2 */
17	-1.09181	/* ST37Q05 step 1 */
18	-0.05693	/* ST37Q05 step 2 */
19	-1.89647	/* ST37Q06 step 1 */
20	0.29021	/* ST37Q06 step 2 */
21	-1.38321	/* ST37Q07 step 1 */
22	0.06433	/* ST37Q07 step 2 */
23	-1.96788	/* ST37Q08 step 1 */
24	0.13774	/* ST37Q08 step 2 */

**Box 6. First six lines of the WLE file matheff.wle**

1	18.00	24.00	1.41200	0.52906
2	13.00	24.00	0.20141	0.47971
3	22.00	24.00	2.80125	0.73009
4	11.00	24.00	-0.24666	0.47732
5	14.00	24.00	0.42920	0.48390
6	9.00	24.00	-0.69620	0.48396

*iii. Instrumental motivation in mathematics*

Four items are used in PFS as well as in PISA 2012 and PISA 2003 to measure instrumental motivation for mathematics (INSTMOT). Table 7 shows the item wording and the international item parameters for this scale. The response categories vary from “Strongly Agree”, “Agree”, “Disagree”, to “Strongly Disagree”. All items must be reversed, so the higher difficulty corresponds to the higher level of motivation. For this index, item difficulties do not vary considerably.

**Table 7. Item parameters for instrumental motivation for mathematics (INSTMOT)**

Item	Thinking about your views on mathematics: to what extent do you agree with the following statements?		Parameter estimates		
			Delta	tau_1	tau_2
ST29Q02	b )	Making an effort in mathematics is worth it because it will help me in the work that I want to do later on	-0.21392	-2.92583	-0.54753
ST29Q05	e )	Learning mathematics is worthwhile for me because it will improve my career <prospects, chances>	-0.26314	-2.80478	-0.65997
ST29Q07	g )	Mathematics is an important subject for me because I need it for what I want to study later on	0.29837	-2.76227	-0.34159
ST29Q08	h )	I will learn many things in mathematics that will help me get a job	0.17869	-2.95336	-0.51783

Source: OECD (2014), *PISA 2012 Technical Report*, Table 16.10, p. 322.

#### iv. Science self-efficacy

Eight items measuring students' science self-efficacy (their confidence in performing science-related tasks) are included in PISA 2006 and remain in PISA 2015. These items cover important themes identified in the science literacy framework: identifying scientific questions, explaining phenomena scientifically and using scientific evidence. Table 8 shows the item wording and the international item parameters for this scale. All items must be reverse coded for IRT scaling so that positive WLE scores on this new index indicate higher levels of self-efficacy in science.

**Table 8. Item parameters for science self-efficacy (SCIEEFF)**

Item	How easy do you think it would be for you to perform the following tasks on your own?		Parameter estimates			
			Delta	tau_1	tau_2	tau_3
ST17Q01	a)	Recognise the science question that underlies a newspaper report on a health issue	-0.28	-1.93	-0.49	2.41
ST17Q02	b)	Explain why earthquakes occur more frequently in some areas than in others	-0.63	-1.49	-0.15	1.64
ST17Q03	c)	Describe the role of antibiotics in the treatment of disease	0.17	-1.55	-0.12	1.68
ST17Q04	d)	Identify the science question associated with the disposal of garbage	0.09	-1.80	-0.21	2.02
ST17Q05	e)	Predict how changes to an environment will affect the survival of certain species	-0.05	-1.48	-0.19	1.67
ST17Q06	f)	Interpret the scientific information provided on the labelling of food items	-0.05	-1.61	-0.17	1.78
ST17Q07	g)	Discuss how new evidence can lead you to change your understanding about the possibility of life on Mars	0.49	-1.43	-0.14	1.57
ST17Q08	h)	Identify the better of two explanations for the formation of acid rain	0.25	-1.46	-0.16	1.62

Source: OECD (2009), *PISA 2006 Technical Report*, Table 16.16, p. 322.

v. *Instrumental motivation in science*

Five items measuring the construct of instrumental motivation are included from PISA 2006 and used to study the trends between PISA 2006 and PISA 2015. Table 9 shows the item wording and the international item parameters for this scale. All items are inverted for IRT scaling: positive WLE scores on this new index indicate higher levels of instrumental motivation to learn science. In PISA 2015, there are 4 items instead of 5 (item ST35Q03 was dropped in PISA 2015). However, the transformation equation links the 5 items from PISA 2006 to the 4 items from PISA 2015. Therefore, instrumental motivation in science can be computed with 5 items, using the parameters from PISA 2006.

**Table 9. Item parameters for instrumental motivation to learn science (INSTSCIE)**

Item	How much do you agree with the statements below? (Strongly agree/Agree/Disagree/Strongly disagree)	Parameter estimates			
		Delta	tau_1	tau_2	tau_3
ST35Q01	a) Making an effort in my <school science> subject(s) is worth it because this will help me in the work I want to do later on	–0.21	–3.46	–0.39	3,85
ST35Q02	b) What I learn in my <school science> subject(s) is important for me because I need this for what I want to study later on	0,24	–3.62	–0.17	3,79
ST35Q03	c) I study <school science> because I know it is useful for me	–0.37	–3.66	–0.67	4,33
ST35Q04	d) Studying my <school science> subject(s) is worthwhile for me because what I learn will improve my career prospects	0,00	–3.66	–0.45	4,11
ST35Q05	e) I will learn many things in my <school science> subject(s) that will help me get a job	0,34	–3.76	–0.29	4,05

*Note:* Item categories were “Strongly agree”, “Agree”, “Disagree” and “Strongly disagree”; all items were inverted for scaling.

*Source:* OECD (2009), *PISA 2006 Technical Report*, Table 16.12, p. 320.

vi. *Disciplinary climate in language of instruction*

This scale provides information on disciplinary climate in the classroom (DISCLIMA) included from PISA 2009. There are five items in this scale, each with four response categories varying from “Strongly Disagree”, “Disagree”, “Agree” to “Strongly Agree”. The items in this scale must be reverse coded (i.e., higher WLE’s on this scale indicate a better disciplinary climate and lower WLE’s a poorer disciplinary climate). Similarly, positive item difficulties indicate aspects of disciplinary climate that are less likely to be found in the classroom environment. Table 10 shows the item wording and the international item parameters for this scale. The item difficulties (deltas) for all the items in this scale are all negative, which means that the items are relatively easier to endorse.

**Table 10. Item parameters for disciplinary climate in language of instruction (DISCLIMA)**

Item	How often do these things happen in your <test language lessons>?	Parameter estimates			
		Delta	tau_1	tau_2	tau_3
ST36Q01	Students don't listen to what the teacher says	-1.9113	-2.7887	-0.5737	3.3623
ST36Q02	There is noise and disorder	-1.8010	-2.6180	-0.3230	2.9410
ST36Q03	The teacher has to wait a long time for students to <quieten down>	-2.0953	-2.3997	-0.3987	2.7983
ST36Q04	Students cannot work well	-2.3637	-3.0073	-0.1563	3.1637
ST36Q05	Students don't start working for a long time after the lesson begins	-2.1123	-2.5987	-0.1797	2.7783

Source: OECD (2012), *PISA 2009 Technical Report*, Table 16.20, p. 296.

### vii. *Disciplinary climate in mathematics*

This scale provides information on disciplinary climate in the mathematics lessons (DISCLIM in PFS but named DISCLIMA in PISA 2012) based on five items. The four response categories are “Every lesson”, “Most lessons”, “Some lessons”, to “Never or hardly ever”. Table 11 shows the item wording and the international item parameters for this scale.

**Table 11. Item parameters for disciplinary climate in mathematics (DISCLIM, PISA 2012)**

Item		How often do these things happen in your mathematics lessons?	Parameter estimates		
			Delta	tau_1	tau_2
ST81Q01	a )	Students don't listen to what the teacher says	0.35916	-2.06346	-0.61398
ST81Q02	b )	There is noise and disorder	0.19734	-1.73779	-0.50655
ST81Q03	c )	The teacher has to wait a long time for students to <quiet down>	-0.09943	-1.54240	-0.41861
ST81Q04	d )	Students cannot work well	-0.30906	-1.71494	-0.55403
ST81Q05	e )	Students don't start working for a long time after the lesson begins	-0.14801	-1.44841	-0.40606

Source: OECD (2014), *PISA 2012 Technical Report*, Table 16.32, p. 331.

### viii. *Disciplinary climate in science*

This scale provides information on the disciplinary climate in science lessons (DISCLICI) based on five items. The four response categories are “Every lesson”, “Most lessons”, “Some lessons” and “Never or hardly ever”. Table 12 shows the item wording and the international item parameters for this scale. Note that these parameters were obtained from a Generalised Partial Credit Model (see Text Box 7 for further explanation).

**Table 12. Item parameters for disciplinary climate in science (DISCLICI, PISA 2015)**

Item	How often do these things happen in your science lessons?		Parameter estimates				
			beta	d_1	d_2	d_3	alpha
ST097Q01	a )	Students don't listen to what the teacher says	0.19029	1.25309	0.51737	-1.77046	0.94803
ST097Q02	b )	There is noise and disorder	0.19407	1.22680	0.34986	-1.57666	1.29726
ST097Q03	c )	The teacher has to wait a long time for students to <quiet down>	-0.00888	1.07093	0.31662	-1.38755	1.14809
ST097Q04	d )	Students cannot work well	-0.33810	1.08205	0.48490	-1.56696	0.79547
ST097Q05	e )	Students don't start working for a long time after the lesson begins	-0.18866	0.99587	0.37880	-1.37468	0.81114

Source: OECD (2017), *PISA 2015 Technical Report*, Table 16.c.

### *ix. Teacher-student relationship*

Five items on teacher-student relations are included from the PISA 2012 Student Questionnaire. This scale provides information on students' perceived teacher's interest in student performance. There are four response categories varying from "Strongly Agree", "Agree", "Disagree" to "Strongly Disagree". All items must be reversed. Table 13 shows the item wording and the international item parameters for this scale. The statement that students generally find the most difficult to agree with is that most of their teachers really listen to what students have to say.

**Table 13. Item parameters for teachers student relations (STUDREL)**

Item	Thinking about the teachers at your school: to what extent do you agree with the following statements?		Parameter estimates		
			Delta	tau_1	tau_2
ST86Q01	a )	Students get along well with most teachers	-0.25440	-2.61793	-0.67095
ST86Q02	b )	Most teachers are interested in students' well-being	0.02701	-2.58166	-0.58553
ST86Q03	c )	Most of my teachers really listen to what I have to say	0.26661	-2.63891	-0.47193
ST86Q04	d )	If I need extra help, I will receive it from my teachers	-0.11302	-2.23629	-0.77270
ST86Q05	e )	Most of my teachers treat me fairly	0.07380	-2.02169	-0.91774

Source: OECD (2014), *PISA 2012 Technical Report*, Table 16.36, p. 333.

### *d. Other complex questionnaire variables*

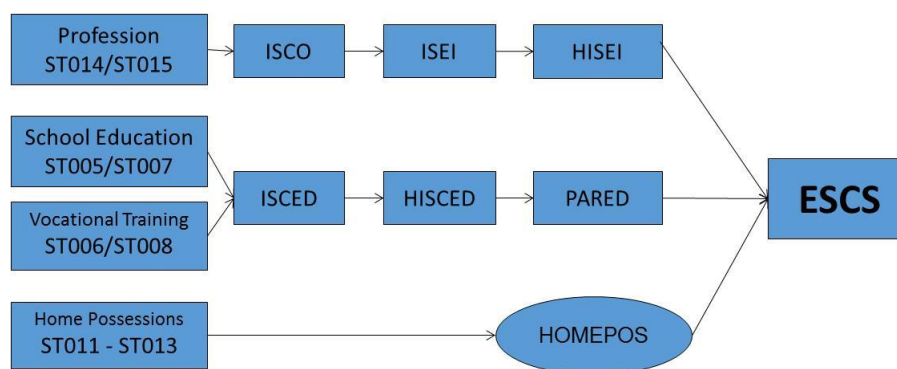
#### *i. Index of economic, social and cultural status (ESCS)*

One key variable for the analyses reported in the school report is the index of economic, social and cultural status (ESCS). The ESCS index was used first in the PISA 2000 analysis, and at that time, was derived from five indices: highest occupational status of parents (HISEI), highest educational level of parents (PARED), and three IRT scales based on student reports on home possessions: family wealth (WEALTH), cultural



possessions (CULTPOSS) and home educational resources (HEDRES). Since PISA 2003, the ESCS has been derived from three indices: highest parental occupation (HISEI), highest educational level of parents (PARED), and one IRT scale based on student reports on home possessions, including books in the home (HOMEPOS). However, until PISA 2012, the principal component analysis (PCA) was based on OECD countries only. In PISA 2015, the PCA is estimated across all countries concurrently. Thus, all countries and economies contribute equally to the estimation of ESCS scores.

**Figure 2. Computation of ESCS in PISA 2015**



Missing values for students with missing data for only one variable must be imputed with predicted values plus a random component based on a regression on the other two variables. If there are missing data on more than one variable, ESCS is not computed for that case and a missing value is assigned for ESCS. The imputed variables were standardised for OECD countries and partner countries/economies with an OECD mean of 0 and a standard deviation of 1.

Standardised variables with imputed values were used in the PCA to obtain ESCS values as component scores for the first principal component across equally weighted countries (OECD and partner countries/economies). It should be noted that in previous cycles, the PCA was based on OECD countries only. For partner countries/economies, ESCS scores were simple indices using standardised imputed variables, fixed factor scores from PCA across OECD countries, and the eigenvalue of the first principal component (see PISA 2012 Technical Report). In PISA 2015, the PCA is estimated across all OECD and partner countries/economies concurrently. Thus, all countries and economies contribute equally to the estimation of ESCS scores. However, for the purpose of reporting, the ESCS scale has been transformed, with 0 being the score of an average

OECD student and 1 being the standard deviation across equally weighted OECD countries.

In the PFS, ESCS scores are obtained as:

$$\text{ESCS} = \frac{\beta_1 \text{HISEI}' + \beta_2 \text{PARED}' + \beta_3 \text{HOMEPOS}'}{\varepsilon_f}$$

where  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  are the OECD factor loadings,  $\varepsilon_f$  is the eigenvalue of the first principal component, HISEI' PARED' and HOMEPOS' are the “OECD-standardised” variables, and  $\varepsilon_f$  equals 1.915. HISEI must be standardised with a mean of 51.50 and a standard deviation of 21.98, PARED with a mean of 13.85 and a standard deviation of 3.08, and HOMEPOS with a mean of 0.00 and a standard deviation of 1.00.

Table 14 shows the factor loadings for the PISA 2015 participating countries and economies.

**Table 14. Factor loadings for PISA 2015 participating countries and economies**

	HISEI	PARED	HOMEPOS	Reliability
Albania	0.82	0.82	0.74	0.69
United Arab Emirates	0.74	0.79	0.48	0.36
Argentina	0.84	0.82	0.75	0.72
Bulgaria	0.82	0.81	0.69	0.67
Brazil	0.82	0.80	0.78	0.71
Colombia	0.81	0.78	0.80	0.70
Costa Rica	0.82	0.79	0.82	0.73
Dominican Republic	0.79	0.77	0.75	0.66
Algeria	0.79	0.76	0.72	0.62
Georgia	0.78	0.76	0.76	0.62
Hong Kong	0.84	0.81	0.76	0.73
Croatia	0.82	0.80	0.70	0.67
Indonesia	0.83	0.81	0.80	0.74
Jordan	0.81	0.83	0.73	0.67
Kazakhstan	0.72	0.77	0.69	0.44
Kosovo	0.76	0.76	0.70	0.58
Lebanon	0.60	0.79	0.77	0.54
Lithuania	0.83	0.81	0.71	0.68
Macao	0.79	0.80	0.70	0.64
Moldova	0.80	0.76	0.76	0.65
Macedonia	0.77	0.76	0.72	0.61
Malta	0.84	0.82	0.65	0.67
Montenegro	0.79	0.76	0.70	0.61

	HISEI	PARED	HOMEPOS	Reliability
Malaysia	0.85	0.76	0.80	0.73
Peru	0.86	0.82	0.81	0.76
Qatar	0.74	0.78	0.50	0.38
B-S-J-G (China)*	0.84	0.80	0.80	0.74
Cyprus	0.85	0.82	0.70	0.70
Romania	0.81	0.78	0.74	0.67
Russian Federation	0.80	0.80	0.70	0.63
Singapore	0.83	0.82	0.77	0.73
Chinese Taipei	0.79	0.80	0.75	0.67
Thailand	0.82	0.81	0.79	0.72
Trinidad and Tobago	0.76	0.75	0.70	0.57
Tunisia	0.83	0.79	0.80	0.73
Uruguay	0.83	0.82	0.77	0.73
Vietnam	0.82	0.83	0.80	0.74
Australia	0.80	0.79	0.67	0.60
Austria	0.81	0.79	0.72	0.66
Belgium	0.84	0.79	0.71	0.68
Canada	0.80	0.79	0.64	0.58
Switzerland	0.82	0.81	0.69	0.68
Chile	0.85	0.84	0.77	0.76
Czech Republic	0.82	0.76	0.72	0.65
Germany	0.83	0.81	0.74	0.70
Denmark	0.83	0.79	0.68	0.65
Spain	0.85	0.83	0.74	0.73
Estonia	0.83	0.78	0.68	0.63
Finland	0.80	0.76	0.68	0.59
France	0.83	0.78	0.72	0.66
United Kingdom	0.80	0.76	0.73	0.63
Greece	0.83	0.82	0.71	0.70
Hungary	0.85	0.83	0.75	0.74
Ireland	0.81	0.80	0.70	0.65
Iceland	0.75	0.76	0.65	0.53
Israel	0.80	0.79	0.68	0.60
Italy	0.83	0.79	0.72	0.68
Japan	0.74	0.76	0.68	0.54
Korea	0.78	0.79	0.73	0.62
Luxembourg	0.86	0.79	0.75	0.72
Latvia	0.83	0.82	0.72	0.69

	HISEI	PARED	HOMEPOS	Reliability
Mexico	0.85	0.85	0.80	0.77
Netherlands	0.81	0.78	0.75	0.67
Norway	0.80	0.78	0.68	0.60
New Zealand	0.81	0.75	0.68	0.58
Poland	0.81	0.80	0.71	0.65
Portugal	0.86	0.84	0.76	0.75
Slovak Republic	0.84	0.82	0.74	0.72
Slovenia	0.84	0.82	0.69	0.68
Sweden	0.82	0.77	0.66	0.61
Turkey	0.82	0.79	0.77	0.68
United States	0.84	0.81	0.74	0.71
Puerto Rico	0.82	0.77	0.72	0.65
Spain	0.84	0.83	0.74	0.72
QMA	0.83	0.82	0.74	0.69
QNC	0.81	0.81	0.74	0.67
* B-S-J-G (China) refers to the four PISA participating china provinces: Beijing, Shanghai, Jiangsu, Guangdong.				

Source: OECD (2017), *PISA 2015 Technical Report*, Chapter 16.

## ii. Highest occupational status of parents (HISEI)

Occupational data for both the student's father and mother are obtained by asking open-ended questions in the Student Questionnaire. The responses must be coded to four-digit International Standard Classification of Occupation (ISCO) codes (ILO, 2007). Details about ISCO codes can be found in the PISA 2012 Student Questionnaire Codebook that will be provided by the OECD. Once the parents' occupations are coded, they must then be mapped to the international socio-economic index of occupational status (ISEI) (Ganzeboom, 2010). In PISA 2015, the new ISCO and ISEI in their 2008 version were used, rather than the 1988 versions that had been applied in the previous four cycles. Like in PISA 2012 and PISA 2015, the new ISCO and ISEI in their 2008 version must be used in the PFS. SAS® and SPSS® programmes for mapping ISCO-08 Occupation code to ISEI-08 are provided by the OECD to the National Service Provider.

Three indices are calculated based on this information: father's occupational status (BFMJ2); mother's occupational status (BMMJ1); and the highest occupational status of parents (HISEI), which corresponds to the higher ISEI score of either parent or to the only available parent's ISEI score. For all three indices, higher scores indicate higher levels of occupational status.

iii. *Education levels of parents (PARED)*

Students' responses regarding parental education must be classified using ISCED (OECD, 1999). Indices on parental education must be constructed by recoding educational qualifications into the following categories: (0) None, (1) ISCED 1 (primary education), (2) ISCED 2 (lower secondary), (3) ISCED Level 3B or 3C (vocational/pre-vocational upper secondary), (4) ISCED 3A (general upper secondary) and/or ISCED 4 (non-tertiary post-secondary), (5) ISCED 5B (vocational tertiary) and (6) ISCED 5A or 6 (theoretically oriented tertiary and post-graduate). Indices with these categories must be created for the students' mother (MISCED) and the students' father (FISCED). In addition, the index on the highest educational level of parents (HISCED) corresponds to the higher ISCED level of either parent. The index for highest educational level of parents must also be recoded into estimated number of years of schooling (PARED). The mapping of ISCED levels to years of schooling (PARED) was updated in 2009 and 2015 for some countries, taking into account changes in countries' educational systems. A mapping of ISCED levels of years of schooling is provided in Table 15.

**Table 15. Mapping of ISCED to years**

Country name	Completed ISCED level 0 (pre-primary education)	Completed ISCED level 1 (primary education)	Completed ISCED level 2 (lower secondary education)	Completed ISCED level 3B or 3C (upper secondary education providing direct access to the labour market or to ISCED 5B programmes)	Completed ISCED level 3A (upper secondary education providing access to ISCED 5A and 5B programmes) and/or ISCED level 4 (non-tertiary post-secondary)	Completed ISCED level 5B (non-university tertiary education)	Completed ISCED level 5A (university level tertiary education) or ISCED level 6 (advanced research programmes)
<b>OECD</b>							
<b>Australia</b>	3	6	10	11	12	14	15
<b>Austria</b>	3	4	9	12	12.5	15	17
<b>Belgium</b>	3	6	9	12	12	15	17
<b>Canada</b>	3	6	9	12	12	15	17
<b>Chile</b>	3	6	8	12	12	16	17
<b>Czech Republic</b>	3	5	9	11	13	16	16
<b>Denmark</b>	3	7	10	13	13	16	18
<b>Estonia</b>	3	6	9	12	12	15	16
<b>Finland</b>	3	6	9	12	12	14.5	16.5

France	3	5	9	12	12	14	15
Germany	3	4	10	13	13	15	18
Greece	3	6	9	11.5	12	15	17
Hungary	3	4	8	10.5	12	13.5	16.5
Country name	Completed ISCED level 0	Completed ISCED level 1	Completed ISCED level 2	Completed ISCED level 3B or 3C	Completed ISCED level 3A	Completed ISCED level 5B	Completed ISCED level 5A
Iceland	3	7	10	13	14	16	18
Ireland	3	6	9	12	12	14	16
Israel	3	6	9	12	12	15	15
Italy	3	5	8	12	13	16	17
Japan	3	6	9	12	12	14	16
Korea	3	6	9	12	12	14	16
Latvia	3	4	8	11	11	14	16
Luxembourg	3	6	9	12	13	16	17
Mexico	3	6	9	12	12	14	16
Netherlands	3	6	10	13	12	15	16
New Zealand	3	5.5	10	11	12	14	15
Norway	3	6	9	12	12	14	16
Poland	3	.	8	11	12	15	16
Portugal	3	6	9	12	12	15	17
Slovak Republic	3	4	9	12	13	16	18
Slovenia	3	4	8	11	12	15	16
Spain	3	5	8	10	12	13	16.5
Sweden	3	6	9	11.5	12	14	16
Switzerland	3	6	9	12.5	12.5	14.5	17.5
Turkey	3	4	8	12	12	14	16
United Kingdom	3	7	10	11	12	14	16
United States	3	6	9	.	12	14	16
<b>Partners</b>							
Albania	3	6	10	12	12	16	16
Algeria	3	5	9	11	12	12	15
Argentina	3	6	10	12	12	14.5	17
Brazil	3	5	9	12	12	14.5	17
B-S-J-G (China)	3	6	9	12	12	15	16.5
Bulgaria	3	4	8	10	12	15	17.5
Colombia	3	5	9	11	11	14	15.5
Costa Rica	3	6	9	11	12	14	16

Croatia	3	4	8	11	12	15	17
*Cyprus	3	6	9	12	12	15	16.5
Dominican Republic	3	6	9	11	12	14	16
Country name	Completed ISCED level 0	Completed ISCED level 1	Completed ISCED level 2	Completed ISCED level 3B or 3C	Completed ISCED level 3A	Completed ISCED level 5B	Completed ISCED level 5A
FYROM	3	5	9	13	13	15	17
Georgia	3	6	9	11	12	13	15.5
Hong Kong (China)	3	6	9	11	13	14	16
Indonesia	3	6	9	12	12	14	15
Jordan	3	6	10	12	12	14.5	16
Kazakhstan	3	4	9	11.5	12.5	14	15
Kosovo	3	5	9	12	14	16	18
Lebanon	3	6	9	12	12	15	16
Lithuania	3	3	8	11	11	15	16
Macao (China)	3	6	9	11	12	15	16
Malaysia	3	6	9	11	13	16	18
Malta	3	6	9	12	13	15	17
Moldova	3	4	9	11	12	14	16.5
Montenegro	3	4	8	11	12	15	16
Peru	3	6	9	11	11	14	17
Qatar	3	6	9	12	12	15	16
Romania	3	4	8	11.5	12.5	14	16
Russia	3	4	9	11	11	13	16
Singapore	3	6	8	10	11	13	16
Chinese Taipei	3	6	9	12	12	14	16
Thailand	3	6	9	12	12	14	16
Trinidad and Tobago	3	5	9	12	12	15	16
Tunisia	3	6	9	12	13	16	17
United Arab Emirates	3	5	9	12	12	15	16
Uruguay	3	6	9	12	12	15	17
Viet Nam	3	5	9	12	12	.	17

\*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.

Source: OECD (2017), *PISA 2015 Technical Report*, Annex D.

#### iv. Household possessions

In PISA 2015, students reported the availability of 16 household items at home (ST011), including three country-specific household items that were seen as appropriate measures of family wealth within the context of the country. In addition, students reported the amount of possessions and books at home (ST012, ST013). HOMEPOS is a summary index of all household and possessions (ST011, ST012 and ST013). HOMEPOS is also one of three components in the construction of the PISA index of economic, social and cultural status. In PFS, the 16 household items are included in the assessment.

The computation of the home possessions scale for PISA 2015 was performed in a way that differed from previous cycles. The IRT model has changed from the Partial Credit model to the Generalised Partial Credit model for the purpose of cross-cultural comparability (see Text Box 7 for further explanation). Categories for the number of books in the home are unchanged in PISA 2015. The variable indicating the number of books at home is recoded from the original 6 categories into 3: (0) 0-25 books, (1) 26-100 books, (2) more than 100. Questions 15-ST012Q01 and 15-ST012Q02 are recoded from the original 4 categories into 3: (0) “None or one”, (1) “Two” (2) “Three or more”. The rest of the questions retain four categories. The ST011 items (1=“yes”, 2=“no”) were reverse-coded, so that a higher level indicates the presence of the indicator.

#### Box 7. Scaling procedures of Context Questionnaire in PISA 2015

As in previous cycles of PISA, one subset of the derived variables was constructed using IRT (item response theory) scaling methodology. In the IRT framework, a number of different models can be distinguished, using the Generalised Partial Credit Model (GPCM) (see below) for constructing derived variables in the PISA 2015 Context Questionnaires.

For each item, item responses are modelled as a function of the latent construct,  $\theta_j$ . With the one-parameter model (Rasch model; Rasch, 1960) for dichotomous items, the probability of person  $j$  selecting category 1 instead of 0 is modelled as

$$P(X_{ji} = 1 | \theta_j, \beta_i) = \frac{\exp(\theta_j - \beta_i)}{1 + \exp(\theta_j - \beta_i)}$$

where  $P(X_{ji}=1)$  is the probability of person  $j$  to score 1 on item  $i$ ,  $\theta_j$  is the estimated latent



trait of person  $j$  and  $\beta_i$  the estimated location or difficulty of item  $i$  on this dimension. In the case of items with more than two ( $m$ ) categories (e.g. Likert-type items), this model can be generalised to the Partial Credit Model (Masters and Wright, 1997), which takes the form of

$$P(X_{ji} = k | \theta_j, \beta_i, d_i) = \frac{\exp(\sum_{u=0}^x \theta_j - \beta_i + d_{ir})}{\sum_{u=0}^{m_i} \exp(\sum_{r=1}^u \theta_j - \beta_i + d_{ir})}$$

where  $P(X_{ji}=k)$  denotes the probability of person  $j$  to score  $k$  on item  $i$  out of the  $m_i$  possible scores on the item.  $\theta_j$  denotes the person's latent trait, the item parameter  $\beta_i$  gives the general location or difficulty of the item on the latent continuum and  $d_{ir}$  denote additional step parameters. This model has been used throughout previous cycles of PISA for scaling derived variables of the context questionnaires. However, research literature (especially, Glass and Jehangir, 2014) suggests that a generalisation of this model, the GCPM (Muraki, 1992), is more appropriate in the context of PISA, since it allows for the item discrimination to vary between items within any given scale. This model takes the form of

$$P(X_{pi} = k | \theta_j, \beta_i, \alpha_i, d_i) = \frac{\exp(\sum_{u=0}^x \alpha_i (\theta_j - \beta_i + d_{ir}))}{\sum_{u=0}^{m_i} \exp(\sum_{r=1}^u \alpha_i (\theta_j - \beta_i + d_{ir}))}$$

in which the additional discrimination parameter  $\alpha_i$  allows for the items of a scale to contribute with different weights to the measurement of the latent construct.

Most of the scales were analysed based on 2015 data only (regular scales) and others, mostly science-related scales were analysed to allow for comparisons with the Weighted Likelihood Estimates (WLE; Warm, 1989) obtained in PISA 2006 (trend scales, see below).

The GCPM described above contains three kinds of item parameters: one relating to the general location or difficulty of the item ( $\beta$ ), one relating to the deviance of each of the single response categories from this location parameter ( $d$ ), and one relating to the item's discrimination or slope ( $\alpha$ ). The following figure displays the category characteristic curves (CCC) of a four-category item (e.g. a Likert-type item with response categories “Strongly disagree”, “Disagree”, “Agree”, and “Strongly agree”). The three kinds of GPCM item parameters were included in this representation. The overall item location or difficulty parameter,  $\beta$ , can be regarded as the item's location on the latent continuum of the construct to be measured. The  $m-1$  threshold parameters,  $d$ , of an  $m$ -category item represent deviations from this general location.

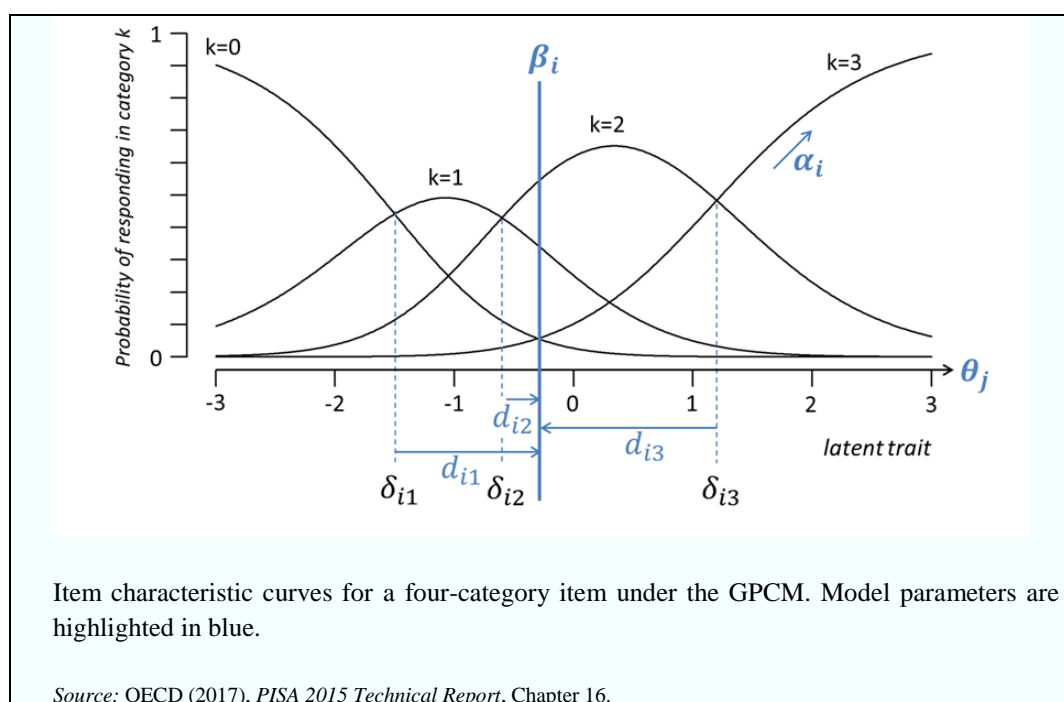


Table 16 shows the wording of items used for the computation of HOMEPOS. Table 17 provides international item parameters for HOMEPOS items ST011, ST012 and ST013. Table 18 provides national item parameters for items ST011Q17 to ST011Q19 (which were nationally scaled).<sup>1</sup>

HOMEPOS is also one of three components in the construction of the index of economic, social and cultural status (or ESCS; see the subsection on ESCS index construction above in this section). The National Service Provider may want to consider recoding the variables using 0 as the minimum value, since it is the default value in CONQUEST, or the syntax may give some errors during the computation. If other software is used, checking the minimum value established as the default is recommended.

**Table 16. Home possession items**

ST011	In your home, do you have:
ST011Q01TA	A desk to study at
ST011Q02TA	A room of your own
ST011Q03TA	A quiet place to study
ST011Q04TA	A computer you can use for school work
ST011Q05TA	Educational software
ST011Q06TA	A link to the Internet
ST011Q07TA	Classic literature (e.g. <Shakespeare>)
ST011Q08TA	Books of poetry
ST011Q09TA	Works of art (e.g. paintings)
ST011Q10TA	Books to help with your school work
ST011Q11TA	<Technical reference books>

ST011Q12TA	A dictionary
ST011Q16NA	Books on art, music, or design
ST011Q17TA	<Country-specific wealth item 1>
ST011Q18TA	<Country-specific wealth item 2>
ST011Q19TA	<Country-specific wealth item 3>
ST12	How many of these are at you home?
ST012Q01TA	Televisions
ST012Q02TA	Cars
ST012Q03TA	Rooms with a bath or shower
ST012Q05NA	<Cell phones> with Internet access (e.g. smartphones)
ST012Q06NA	Computers (desktop computer, portable laptop, or notebook)
ST012Q07NA	<Tablet computers> (e.g. <iPad®>, <BlackBerry® PlayBook™>)
ST012Q08NA	E-book readers (e.g. <Kindle™>, <Kobo>, <Bookeen>)
ST012Q09NA	Musical instruments (e.g., guitar, piano)
ST013Q01TA	How many books are there in your home?

Source: OECD (2017), *PISA 2015 Technical Report*, Chapter 16.

**Table 17. International item parameters for HOMEPOS**

Item	Parameter estimates						
	beta	d_1	d_2	d_3	d_4	d_5	alpha
ST011Q01TA A desk to study at	-0.996						0.996
ST011Q02TA A room of your own	-0.815						0.767
ST011Q03TA A quiet place to study	-1.137						0.813
ST011Q04TA A computer you can use for school work	-0.345						2.030
ST011Q05TA Educational software	0.340						0.952
ST011Q06TA A link to the Internet	-0.417						2.448
ST011Q07TA Classic literature (e.g. <Shakespeare>)	0.229						0.590
ST011Q08TA Books of poetry	-1.387						0.485
ST011Q09TA Works of art (e.g. paintings)	0.151						0.870
ST011Q10TA Books to help with your school work	-1.226						0.593
ST011Q11TA <Technical reference books>	0.188						0.886
ST011Q12TA A dictionary	-1.746						0.701
ST011Q16NA A Books on art, music, or design	-1.027						1.256
ST011Q17TA <Country-specific wealth item 1>	-0.898						1.362
ST011Q18TA <Country-specific wealth item 2>	-0.723						1.465
ST011Q19TA <Country-specific wealth item 3>	-1.174						1.004
ST012Q01TA Televisions	-0.740	1.905	-0.718	-1.187			0.623

ST012Q02TA	Cars	0.562	0.744	-0.056	-0.688	0.979		
ST012Q03TA	Rooms with a bath or shower	0.437	1.356	-0.416	-0.939	0.982		
ST012Q05N A	<Cell phones> with Internet access (e.g., smartphones)	-0.452	0.362	-0.507	0.145	0.838		
ST012Q06N A	Computers (desktop computer, portable laptop, or notebook)	0.206	0.632	-0.169	-0.464	1.691		
ST012Q07N A	<Tablet computers> (e.g. <iPad®>, <BlackBerry® PlayBook™>)	0.812	0.487	-0.305	-0.182	0.876		
ST012Q08N A	E-book readers (e.g. <Kindle™>, <Kobo>, <Bookeen>)	1.796	-0.241	-0.254	0.495	0.647		
ST012Q09N A	Musical instruments (e.g. guitar, piano)	0.883	0.125	-0.308	0.183	0.651		
ST013Q01TA	How many books are there in your home?	0.840	0.679	0.829	-0.541	-0.286	-0.680	0.494

Source: OECD (2017), *PISA 2015 Technical Report*, Chapter 16.

Table 18. National item parameters for HOMEPOS

Country name	ST011Q17TA		ST011Q18TA		ST011Q19TA	
	beta	alpha	beta	alpha	beta	alpha
<b>OECD</b>						
Australia	1.638	0.717	1.039	0.710	0.970	1.200
Austria	-0.136	1.341	-0.730	0.674	1.187	1.250
Belgium (Flemish)	0.606	0.864	0.890	1.159	0.844	1.801
Belgium (French)	-0.238	0.720	0.505	0.812	0.902	1.712
Canada	-0.715	1.250	0.858	0.869	-0.041	0.845
Chile	0.988	0.749	0.085	1.701	-0.654	1.198
Czech Republic	N/A		N/A		N/A	
Denmark	0.847	2.910	N/A		N/A	
Estonia	0.118	1.448	-0.083	1.901	0.582	1.616
Finland	-0.012	2.205	0.314	0.870	N/A	
France	-0.169	0.918	0.036	1.428	-0.230	1.423
Germany	N/A		1.443	-0.556	0.752	2.359
Greece	0.165	1.687	-0.027	1.118	0.828	1.410
Hungary	0.058	0.718	0.156	1.134	-0.025	2.355
Iceland	0.820	1.245	1.438	1.217	1.375	0.718
Ireland	-0.223	1.138	-0.186	0.812	-1.646	0.831
Israel	1.270	0.865	0.859	1.409	1.066	1.052
Italy	0.610	1.254	0.810	1.134	0.286	1.068
Japan	-0.444	1.934	-2.544	0.724	-0.170	1.010
Korea	-0.195	1.373	0.893	1.488	0.774	1.624
Latvia	-0.616	1.768	-0.371	1.258	2.291	0.497
Luxembourg	-0.645	0.617	0.584	1.317	-1.044	0.329
Mexico	0.235	1.089	-0.377	1.254	-0.479	1.362
Netherlands	0.752	0.783	1.117	2.825	0.204	1.102
New Zealand	0.271	0.893	1.686	0.854	1.603	1.307
Norway	0.084	1.875	-0.360	0.915	N/A	
Poland	0.463	1.802	-0.093	2.369	0.066	2.315
Portugal	-1.156	0.705	-0.393	1.267	1.027	1.201
Slovak Republic	0.053	1.290	-0.044	2.296	N/A	
Slovenia	-0.887	0.604	0.421	1.834	0.326	1.529
Spain	-0.350	1.588	0.348	0.716	1.115	0.867
Sweden	1.099	1.572	1.487	0.826	0.857	0.916
Switzerland	0.532	2.438	-0.736	0.630	0.003	1.416
Turkey	-0.121	1.107	-0.236	1.735	0.902	1.487
United Kingdom (excl. Scotland)	-0.938	0.442	-0.865	0.940	-0.282	1.389

	ST011Q17TA		ST011Q18TA		ST011Q19TA	
<b>United Kingdom (Scotland)</b>	N/A		0.639	1.619	1.900	1.472
<b>United States</b>	0.901	0.871	-0.197	1.392	0.469	1.258
<b>Partners</b>						
<b>Albania</b>	-0.898	1.362	-0.723	1.465	-1.174	1.004
<b>Algeria</b>	N/A		N/A		N/A	
<b>Argentina</b>	0.484	0.346	0.035	1.244	-1.125	1.783
<b>Brazil</b>	0.443	1.403	-0.063	1.230	0.782	1.493
<b>B-S-J-G (China)</b>	0.450	2.111	0.018	2.436	-0.082	1.335
<b>Bulgaria</b>	-0.282	2.199	0.001	2.175	0.236	1.363
<b>Colombia</b>	-0.129	1.721	N/A		-0.877	0.915
<b>Costa Rica</b>	-0.759	1.278	0.034	1.592	0.757	1.225
<b>Croatia</b>	-0.087	0.935	0.209	1.117	0.342	1.535
<b>*Cyprus</b>	1.123	1.046	1.760	1.340	1.031	1.370
<b>Dominican Republic</b>	-0.139	1.330	0.302	1.710	-0.539	1.282
<b>FYROM</b>	0.956	1.212	1.419	0.979	N/A	
<b>Georgia</b>	0.704	1.073	0.933	1.225	1.596	1.228
<b>Hong Kong (China)</b>	-0.090	1.255	0.660	2.111	0.672	0.796
<b>Indonesia</b>	0.017	1.459	-1.723	1.063	0.168	2.101
<b>Jordan</b>	0.219	0.916	0.017	1.281	0.117	1.650
<b>Kazakhstan</b>	-0.006	2.677	0.000	2.047	-0.495	1.169
<b>Kosovo</b>	-1.289	1.342	-0.562	1.459	0.126	1.876
<b>Lebanon</b>	-0.388	1.407	-0.641	1.397	-1.654	0.680
<b>Lithuania</b>	0.053	2.357	0.572	1.141	0.202	1.783
<b>Macao (China)</b>	0.651	2.051	-0.095	2.325	0.206	1.803
<b>Malaysia</b>	-3.237	0.737	-2.647	0.821	0.203	1.550
<b>Malta</b>	1.282	1.241	1.499	0.847	1.809	1.167
<b>Moldova</b>	-0.070	1.910	N/A		N/A	
<b>Montenegro</b>	-0.500	2.070	-0.122	2.063	0.010	2.284
<b>Peru</b>	0.310	1.438	-0.641	1.875	-0.105	2.037
<b>Qatar</b>	0.223	0.897	0.064	1.556	0.020	1.227
<b>Romania</b>	-0.759	1.192	-1.342	0.788	0.068	1.899
<b>Russia</b>	1.183	2.220	0.714	1.667	0.701	1.464
<b>Singapore</b>	-0.049	1.836	1.109	1.471	N/A	
<b>Chinese Taipei</b>	0.560	2.208	-0.232	1.333	-0.187	1.744
<b>Thailand</b>	0.238	2.405	1.625	1.304	0.180	1.981
<b>Trinidad and Tobago</b>	-0.559	1.159	-1.388	0.813	-0.793	0.536
<b>Tunisia</b>	-0.410	1.679	-0.068	1.912	-1.374	1.295
<b>United Arab Emirates</b>	0.185	1.389	0.205	1.297	0.157	1.432

	ST011Q17TA		ST011Q18TA		ST011Q19TA	
<b>Uruguay</b>	-0.088	0.487	-1.176	1.585	-0.072	2.496
<b>Viet Nam</b>	0.087	2.643	-2.239	0.989	0.566	2.121

\*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”.

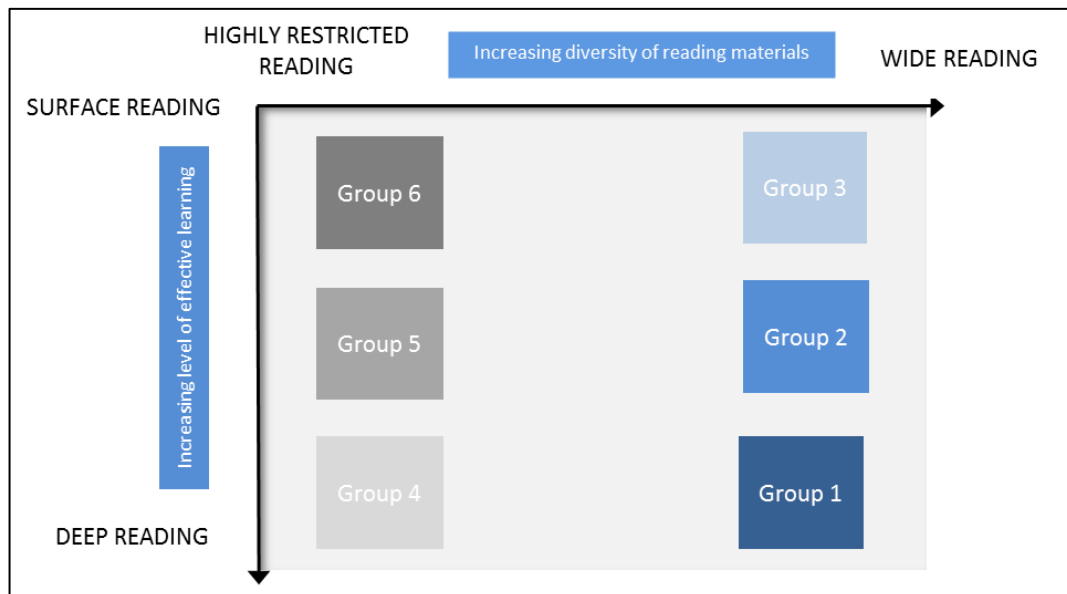
Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of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

Source: OECD (2017), *PISA 2015 Technical Report*, Table 16.

### v. Reader profiles

Students that participated in the PFS are classified into six profiles of readers based on whether they regularly read comic books, magazines, newspapers, fiction and non-fiction books for enjoyment – as an indicator of how “wide” their reading habits are – as well as on their awareness of effective learning strategies to understand, remember and summarise information – as indicators of how “deep” their reading and learning is. Figure 3 illustrates how the reading process can be characterised along the width and depth dimensions.

**Figure 3. How the reading process can be characterised**



Source: OECD (2010), *PISA 2009 Results: Learning to Learn: Student Engagement, Strategies and Practices (Volume III)*, Figure III.1.19.

As in PISA 2009, students in PFS must be grouped into the six profiles shown in Figure 4. Latent class analysis must be performed to classify students into the six profiles.

This subsection first describes how to create the seven variables used in the latent profile analysis and then explains how to perform the latent profile analysis in order to classify students into reader profiles. *Mplus* syntax to perform these analyses will be provided by the OECD.

**Figure 4. Profile of readers**

	<b>Reads all material regularly</b>	<b>Reads magazines and newspapers regularly</b>	<b>Does not read any material regularly</b>
<b>High levels of effective learning strategies</b>	Deep and wide readers	Deep and narrow readers	Deep and highly restricted readers
<b>Low levels of effective learning strategies</b>	Surface and wide readers	Surface and narrow readers	Surface and highly restricted readers

Source: OECD (2010), *PISA 2009 Results: Learning to Learn: Student Engagement, Strategies and Practices (Volume III)*, Figure III.1.20.

### Construction of variables

Questions ST25Q01 to ST25Q05 included from PISA 2009 Student Questionnaire about students' frequency of reading certain materials must be recoded into five dummy variables named "magazine", "comic", "fiction", "nonfic" and "news". Each of these dummy variables equals one if the student declares reading this material "several times a month" or "several times a week" and zero if the student declares reading this material "never or almost never", "a few times a year" or "about once a month".

Two other variables must be derived from questions relative to two meta-cognition tasks "Understanding and remembering" (UNDREM) and "Summarising" (METASUM). Both meta-cognition tasks consist of a stem (which is a reading task) and a set of strategies.

For each strategy listed in items ST41Q01 to ST41Q06 from the PISA 2009 Student Questionnaire, students are asked to rate the usefulness of the strategy for understanding and remembering a text, using a scale ranging from 1 to 6. For each student, nine scores must be calculated according to the following set of nine order relations:

- If ST41Q03 > ST41Q01 then student's score is set to 1, else 0.
- If ST41Q03 > ST41Q02 then student's score is set to 1, else 0.
- If ST41Q03 > ST41Q06 then student's score is set to 1, else 0.
- If ST41Q04 > ST41Q01 then student's score is set to 1, else 0.
- If ST41Q04 > ST41Q02 then student's score is set to 1, else 0.
- If ST41Q04 > ST41Q06 then student's score is set to 1, else 0.
- If ST41Q05 > ST41Q01 then student's score is set to 1, else 0.



- If ST41Q05>ST41Q02 then student's score is set to 1, else 0.
- If ST41Q05>ST41Q06 then student's score is set to 1, else 0.

If any of the rates used in an order relation is missing, then the associated score must also be missing. The nine created scores must be added together and divided by nine. The resulting proportion is the student's score on the UNDREM scale.

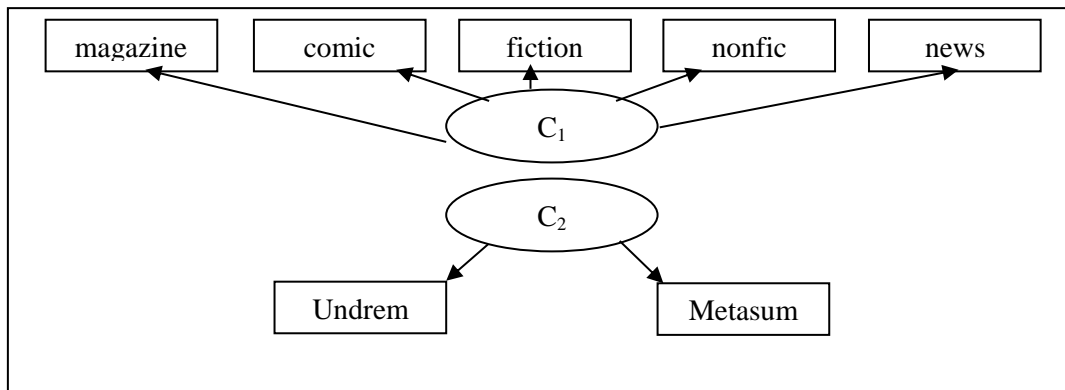
For each strategy listed in items ST42Q01 to ST42Q05 from the PISA 2009 Student Questionnaire, students are asked to rate the usefulness of the strategy for *summarising* a text, using a scale ranging from 1 to 6. For each student, eight scores must be calculated according to the following set of eight order relations:

- If ST42Q04>ST42Q01 then student's score is set to 1, else 0.
- If ST42Q04>ST42Q03 then student's score is set to 1, else 0.
- If ST42Q04>ST42Q02 then student's score is set to 1, else 0.
- If ST42Q05>ST42Q01 then student's score is set to 1, else 0.
- If ST42Q05>ST42Q03 then student's score is set to 1, else 0.
- If ST42Q05>ST42Q02 then student's score is set to 1, else 0.
- If ST42Q01>ST42Q02 then student's score is set to 1, else 0.
- If ST42Q03>ST42Q02 then student's score is set to 1, else 0.

If any of the rates used in an order relation is missing, then the associated score must also be missing. The eight created scores must be added together and divided by eight. The resulting proportion is the student's score on the METASUM scale.

### Latent profile analysis

Latent profile analysis must be performed to identify profile of readers. This model assumes that there are two dimensions (i.e., latent variables) when classifying students into groups. Figure 5 illustrates the two categorical latent variables model (a model with two dimensions) employed to estimate profiles of students. Figure 5 shows how the first dimension, characterising the material students read several times a month or several times a week – i.e., categorical latent variable C1 – is identified by five variables: whether students read regularly comic books, fiction, non-fiction, magazines and newspapers. The second dimension, characterising students' awareness of effective learning strategies – i.e., categorical latent variable C2 – is identified by two variables: awareness of effective strategies to understand and remember information (UNDREM) and awareness of effective strategies summarise information (METASUM). These two meta-cognition indices (UNDREM and METASUM) must be standardised to an OECD mean of 0 and a standard deviation of 1 (for the pooled data with equally weighted country samples). UNDREM must be standardised with a mean of 0.57 and a standard deviation of 0.30, and METASUM with a mean of 0.60 and a standard deviation of 0.29.

**Figure 5. Latent profile analysis with two categorical latent variables**

Means are specified to vary only across the classes within each dimension. First, students are grouped into classes within each dimension, and then groupings are assigned according to the combination of the two dimensions  $C_1$  and  $C_2$ . Three classes must be extracted from the first dimension, and two classes from the second dimension, resulting in a total of 6 groups that students can be assigned to ( $3 \times 2$ ). Students' assignment to reader profiles must be performed using coefficients fixed at their international values. Input and output files of the latent class analysis performed with *Mplus* software on PISA 2009 data will be provided by the OECD.

## 10. Survey weighting (the balanced repeated replication method)

Survey weights are required to facilitate calculation of appropriate estimates of sampling error and making valid estimates and inferences of the population. Statisticians must calculate survey weights for all assessed students, to get estimates of standard errors, conduct significance tests and create confidence intervals appropriately.

A replication methodology must be employed to estimate the sampling variances of PFS parameter estimates. This methodology accounts for the variance in estimates due to the sampling of students. Additional variance due to the use of plausible values from the posterior distributions of scaled scores is captured separately as measurement error. Computationally the calculation of these two components can be carried out using SAS® and SPSS® macros. Additional variance due to the comparison of PFS results with those of PISA is captured separately by a link error. For further detail, see *PISA Data Analysis Manual* (OECD, 2009a or 2009b).

### *a. The balanced repeated replication method*

The approach used for calculating sampling variances for PISA and therefore PFS estimates is known as Balanced Repeated Replication (BRR), or balanced half-samples. The particular variant known as Fay's method must be used.<sup>2</sup> This variant of the BRR method must be implemented as follows.

In each school, students must be ranked on the basis of their first principal component that was used to generate the plausible values. The first two students in the ordered list are paired; the following two students are paired and so on. If a school has an odd number of students, the last three students are grouped to form a triple. Let us assume that there are  $H$  pairs of students, also referred as variance strata or zones, or pseudo-strata in the sampling literature. Pairs are numbered sequentially, 1 to  $H$ .

A set of 80 replicate weights must then be created. Each of these replicate weights is formed by assigning a weight of 1.5 to one of the two students in the stratum, and a weight of 0.5 to the remaining student. In cases where there are three units in a triple, either one of the students (designated at random) receive a factor of 1.7071 for a given replicate, with the other two students receiving factors of 0.6464, or else the one student receives a factor of 0.2929, and the other two students receive factors of 1.3536. The determination as to which students receive inflated weights and which receive deflated weights is carried out in a systematic fashion, based on the entries of the first  $H$  rows in a Hadamard matrix of order 80. A Hadamard matrix<sup>3</sup> contains entries that are +1 and -1 in value, and has the property that the matrix, multiplied by its transpose, gives the identity matrix of order 80, multiplied by a factor of 80.

For pairs of students:

- The +1 in the Hadamard matrix is converted to a weight of 1.5 for the first student of the pair, and 0.5 for the second student of the pair;
- The -1 in the Hadamard matrix is converted to a weight of 0.5 for the first student of the pair, and 1.5 for the second student of the pair.

For triples of students:

- The +1 in the Hadamard matrix is converted to a weight of 1.7071 for the first student of the pair, and 0.6464 for the other two students of the triple;
- The -1 in the Hadamard matrix is converted to a weight of 0.2929 for the first student of the pair, and 1.3536 for the other two students of the triple.

Table 19 and Table 20 describe how the replicate weights are generated for this method in a fictitious example where there are 21 students in a participating school. Table 19 displays an example of a Hadamard matrix. Table 19 shows how the replicate weights are assigned to each student.

Table 19. Hadamard Matrix

	Column 1	Column 2	Column 3	...	Column 80
Row 1	1	1	1		1
Row 2	1	1	-1		-1
Row 3	1	-1	1		1
Row 4	1	1	-1		1
Row 5	1	1	1		1
Row 6	1	1	1		-1
Row 7	1	-1	1		-1
Row 8	1	-1	-1		-1
Row 9	-1	-1	-1		1
Row 10	1	1	-1		-1

Table 20. Replicates for the Balanced Replicate method

Pseudo-statum	Student identifier	FULL WEIGHT	BRR WEIGHT 1	BRR WEIGHT 2	BRR WEIGHT 3	...	BRR WEIGHT 80
1	1	1	1.5	1.5	1.5		1.5
1	2	1	0.5	0.5	0.5		0.5
2	3	1	1.5	1.5	0.5		0.5
2	4	1	0.5	0.5	1.5		1.5
3	5	1	1.5	0.5	1.5		1.5
3	6	1	0.5	1.5	0.5		0.5
4	7	1	1.5	1.5	0.5		1.5
4	8	1	0.5	0.5	1.5		0.5
5	9	1	1.5	1.5	1.5		1.5
5	10	1	0.5	0.5	0.5		0.5
6	11	1	1.5	1.5	1.5		0.5
6	12	1	0.5	0.5	0.5		1.5
7	13	1	1.5	0.5	1.5		0.5
7	14	1	0.5	1.5	0.5		1.5
8	15	1	1.5	0.5	0.5		0.5
8	16	1	0.5	1.5	1.5		1.5
9	17	1	0.5	0.5	0.5		1.5
9	18	1	1.5	1.5	1.5		0.5
10	19	1	0.7071	0.7071	0.2929		0.2929
10	20	1	0.6464	0.6464	1.3536		1.3536
10	21	1	0.6464	0.6464	1.3536		1.3536

Source: Based on OECD (2009), *PISA Data Analysis Manual. Second edition*, Table 4.12/4.13.

### ***b. The sampling variance estimator***

As with all replication methods, the statistic of interest ( $\varphi$ ) is computed based on the sample ( $\widehat{\varphi}$ ), and then again on each replicate ( $\widehat{\varphi}_{(i)}, i = 1$  to 80). The replicates are then compared to the whole sample estimate to get the sampling variance, as follows:

$$\sigma^2_{(\widehat{\varphi})} = \frac{1}{20} \sum_{i=1}^{80} (\widehat{\varphi}_{(i)} - \widehat{\varphi})^2$$

## **11. Statistical procedures for generating school report results**

This section of the report has been developed to provide Statisticians with the techniques needed to correctly analyse the PFS database. It helps them to confidently and accurately implement procedures used for the production of the PFS school reports. This section is largely based on the procedures described in the *PISA Data Analysis Manual* (OECD, 2009a and 2009b) which the reader can refer to for further detail. This section will cover all cases of computations required for generating school report results.

### ***a. Univariate statistics for context variables***

Replicate weights have to be used for the computation of the standard error for any population estimate. The standard error of statistics for context variables only (i.e., not involving performance variables) simply equals the sampling error, i.e., the square root of the sampling variance estimator reported in Section 10.b.

#### ***i. Percentage of students per category***

For categorical variables, the statistic of interest is usually a percentage of students per category. Each percentage is estimated after deleting cases with missing values for the variable of interest in the dataset. The procedure for estimating the percentage of students per category and the corresponding standard error requires first computing the percentage for the original sample (without weighting student data), and then computing 80 other percentages, each of them by weighting the student sample with one of the 80 replicates.

SAS® and SPSS® syntaxes and macros (respectively PROC\_FREQ\_NO\_PV.SAS and MCR\_SE\_GRPPCT.SPS) for computing the percentages and their standard errors are presented in chapter 7 of the *PISA Data Analysis Manual* (OECD, 2009a and 2009b).

For reporting, percentages of students:

- reporting that disciplinary issues in <test language> lessons occur “never or hardly ever” or “in some lessons”;
- reporting that disciplinary issues in mathematics lessons occur “never or hardly ever” or “in some lessons”;

- who agree or strongly agree with statements regarding teacher-student relations at school;
- per reading profile;
- who feel either confident or very confident about having to do various mathematics tasks;
- who feel either confident or very confident about having to do various science tasks;
- who agree or strongly agree with statements regarding their instrumental motivation in mathematics;
- who agree or strongly agree with statements regarding their instrumental motivation in science;

as well as their standard errors must be computed.

#### *ii. School mean indices*

To compute a school mean index and its corresponding standard error, it is also necessary to first compute the mean on the original student sample, and then to compute 80 other means, each of them by weighting the data with one of the 80 replicates.

SAS® and SPSS® syntaxes and macros (respectively PROCMEAN\_NO\_PV.SAS and MCR\_SE\_UNIV.SPS) for computing the percentages and their standard errors are presented in chapter 7 of the *PISA Data Analysis Manual* (OECD, 2009a and 2009b).

For reporting, the school mean of:

- the economic, social and cultural status index (ESCS);
- the index of disciplinary climate in <test language> lessons (DISCLIMA);
- the index of disciplinary climate in mathematics (DISCLIM);
- the index of teacher-student relations (STUDREL);
- the index of instrumental motivation in mathematics (INSTMOT);
- the index of instrumental motivation in science (INSTSCIE);
- the index of self-concept in mathematics (MATHEFF);
- the index of self-concept in science (SCIEEFF);

as well as their standard errors must be computed.

For reporting purposes, the DISCLIM and STUDREL indices (which are originally expressed in a metric with an OECD mean of 0 and an OECD standard deviation of one) must be converted to a scale with an OECD mean of 5 and a standard deviation of 2,

using the following transformation: Transformed index = 2 index + 5, where index is either DISCLIM or STUDREL.

*iii. Identification of “similar” schools in regards to socio-economic background*

Several figures in the school report compare the school’s results with those of similar schools with regards to socioeconomic background. “Similar” schools are schools that participated in PISA 2012 and whose ESCS is within +/- 0.25 of the ESCS index of the school having participated in the PFS.

***b. Students average scores and their standard errors***

As described above, the cognitive data in the PFS are scaled with the mixed coefficients multinomial logit model and the performance of students is denoted with plausible values (PVs). For each domain, five plausible values per student are included in the international databases. This section describes how to perform analyses with plausible values, so it is useful when reporting results on student performances and their relationships with student or school characteristics.

The computation of a statistic with plausible values always consists of six steps, regardless of the required statistic.

1. The required statistic has to be computed for each plausible value. In the previous section, it was mentioned that 81 estimates are necessary to get the final estimate and its standard error. Therefore, any analysis that involves five plausible values will require 405 estimates ( $5 \times 81$ ). To estimate a mean score and its respective standard error, 405 means must be calculated. The means estimated on the original sample (without weighting) are denoted  $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4$  and  $\hat{\mu}_5$ . From the 80 replicates applied on each of the five plausible values, five sampling variances are estimated, denoted respectively  $\sigma^2_{(\hat{\mu}_1)}, \sigma^2_{(\hat{\mu}_2)}, \sigma^2_{(\hat{\mu}_3)}, \sigma^2_{(\hat{\mu}_4)}$  and  $\sigma^2_{(\hat{\mu}_5)}$ . These five mean estimates and their respective sampling variances are provided in Table 21.

**Table 21. The 405 mean estimates**

Weight	PV1	PV2	PV3	PV4	PV5
None	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	$\hat{\mu}_5$
Replicate 1	$\hat{\mu}_{1,1}$	$\hat{\mu}_{2,1}$	$\hat{\mu}_{3,1}$	$\hat{\mu}_{4,1}$	$\hat{\mu}_{5,1}$
Replicate 2	$\hat{\mu}_{1,2}$	$\hat{\mu}_{2,2}$	$\hat{\mu}_{3,2}$	$\hat{\mu}_{4,2}$	$\hat{\mu}_{5,2}$
Replicate 3	$\hat{\mu}_{1,3}$	$\hat{\mu}_{2,3}$	$\hat{\mu}_{3,3}$	$\hat{\mu}_{4,3}$	$\hat{\mu}_{5,3}$
....	...	....	...	...	...
Replicate 80	$\hat{\mu}_{1,80}$	$\hat{\mu}_{2,80}$	$\hat{\mu}_{3,80}$	$\hat{\mu}_{4,80}$	$\hat{\mu}_{5,80}$
Sampling variance	$\sigma^2_{(\hat{\mu}_1)}$	$\sigma^2_{(\hat{\mu}_2)}$	$\sigma^2_{(\hat{\mu}_3)}$	$\sigma^2_{(\hat{\mu}_4)}$	$\sigma^2_{(\hat{\mu}_5)}$

Source: OECD (2009), *PISA Data Analysis Manual*, Second edition, Table 8.1.

2. The final mean estimate is equal to the average of the five mean estimates, i.e.,

$$\hat{\mu} = \frac{1}{5}(\hat{\mu}_1 + \hat{\mu}_2 + \hat{\mu}_3 + \hat{\mu}_4 + \hat{\mu}_5).$$

3. The final sampling variance is equal to the average of the five sampling variances, i.e.,

$$\hat{\sigma}_{\text{sampling}}^2 = \frac{1}{5}(\sigma_{(\hat{\mu}_1)}^2 + \sigma_{(\hat{\mu}_2)}^2 + \sigma_{(\hat{\mu}_3)}^2 + \sigma_{(\hat{\mu}_4)}^2 + \sigma_{(\hat{\mu}_5)}^2).$$

4. The imputation variance, also denoted measurement error variance, is computed as

$$\hat{\sigma}_{\text{measure}}^2 = \frac{1}{4} \sum_{i=1}^5 (\hat{\mu} - \hat{\mu}_i)^2$$

5. Two types of standard errors can be computed depending on the reporting purpose. Whenever a mean score is compared with a specific PISA result or put in parallel of a specific PISA scale, the appropriate link error must be included in the calculation of the standard error of the mean estimate. For instance, if the school average score of the PFS mathematics scale is compared with the country's average score at PISA 2015 mathematics test, the link error between the PFS and the PISA 2015 mathematics tests must be included in the estimation of the standard error on the school's average score. The total error variance results from the combination of the sampling error variance, the measurement error variance and the linking error variance (denoted  $\hat{\sigma}_{\text{link}}^2$ ), and it is computed as:

$$\hat{\sigma}_{\text{final with link error}}^2 = 1.2 \hat{\sigma}_{\text{measure}}^2 + \hat{\sigma}_{\text{sampling}}^2 + \hat{\sigma}_{\text{link}}^2.$$

Whenever a mean estimate (such as the school average score) is presented as such, i.e., alone, then no link error must be taken into account in the calculation of the standard error. The sampling variance and the imputation variance are combined to obtain a final error variance as:

$$\hat{\sigma}_{\text{final without linking}}^2 = 1.2 \hat{\sigma}_{\text{measure}}^2 + \hat{\sigma}_{\text{sampling}}^2.$$

6. The standard error equals the square root of the total error variance.

The SAS® macro PROC\_MEANS\_PV.SAS and the SPSS® macro MCR\_SE\_PV.SPS for computing mean scores and their standard errors (without linking) are described in chapter 8 of the *PISA Data Analysis Manual*. Link errors are given in Table 22.



**Table 22. Link error estimates**

Domain	Link	Link error on PISA point scale
Mathematics	PBTS - PISA 2003	4.082
	PBTS - PISA 2012	5.166
	PBTS - PISA 2015	10.101
Reading	PBTS - PISA 2009	6.093
	PBTS - PISA 2012	7.261
	PBTS - PISA 2015	8.132
Science	PBTS - PISA 2006	6.163
	PBTS - PISA 2012	7.172
	PBTS - PISA 2015	11.188

For reporting, the following scores averages and their standard errors must be computed:

- Average score of students in the school.
- Average score of students in the school, by gender.
- Average score of students in the school, by reader profile.

*iv. School mean performance adjusted for ESCS*

A multilevel regression of students' scores with random intercepts and one fixed effect (student's ESCS centred at the national ESCS average) must be performed to compute school mean performance adjusted for ESCS. The equation of this model can be written as:

$$Y_{ij} = \gamma_{00} + \gamma_{10}(\text{ESCS}_{ij} - \overline{\text{ESCS}}) + u_{0j} + \varepsilon_{ij}$$

where  $Y_{ij}$  represents the performance of student  $i$  in school  $j$ ,  $\gamma_{00}$  the overall intercept,  $\gamma_{10}$  the regression coefficient for centred student's ESCS. The model has two random components:  $u_{0j}$ , the school departure from the overall intercept, and  $\varepsilon_{ij}$  the student residual.

This can be estimated with the SAS® or the SPSS® syntaxes and macros (proc\_mixed\_pv.sas and MCR\_ML\_PV.SPS) detailed in chapter 15 of the *PISA Data Analysis Manual*. The average school performance after accounting for ESCS must be computed as the sum of the fixed intercept ( $\gamma_{00}$ ) and the random component of the intercept ( $u_{0j}$ ) for the considered school.

*v. Distribution of students across proficiency levels*

The SAS® macro proc\_freq\_pv.sas and the SPSS® macro MCR\_SE\_PCTLEV.SPS have been developed for computing the percentage of students at each proficiency level

as well as its respective standard error in one run. They are detailed in chapter 9 of the *PISA Data Analysis Manual*.

***c. Reporting and reviewing school-level data***

National Service Providers are asked to prepare an Excel workbook containing, for all participating schools, all school-level data that will be used as input in the school report. This workbook typically contains school observations in rows and all required school-level variables in columns. National Service Providers usually process altogether the data of schools that participated in the same testing window. So, once the data processing of a given testing window is completed, the National Service Provider must fill in the Excel workbook with the corresponding school data, and send it for review to the OECD. The OECD will perform several quality assurance checks on the dataset and possibly send feedback to the National Service Provider. Once the calculated results have been approved by the OECD, the National Service Provider can proceed with the eBook production for the schools of this testing window. This process will be repeated for each testing window.

## Notes

- <sup>1</sup> For a country that has not participated in PISA yet, two solutions can be envisaged. One is to use national item parameters of another relevant participating country. The other is to use freely estimated item parameters for items ST26Q13 to ST26Q17 when computing the HOMEPOS index.
- <sup>2</sup> This method is similar in nature to the jackknife method used in other international studies of educational achievement.
- <sup>3</sup> Details concerning Hadamard matrices are given in Wolter (2007).

## References

- Adams, R.J., M. Wilson and W.C. Wang (1997), “The Multidimensional Random Coefficients Multinomial Logit Model”, *Applied Psychological Measurement*, No. 21, pp. 1-23.
- Glass, C. A. W., & Jehangir, K. (2014), *Modeling country-specific differential item functioning*, In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis*. (pp. 97-115), Springer, New York.
- Ganzeboom, H.B.G. (2010), “Questions and answers about ISEI-08”, [www.harryganzeboom.nl/isco08/qa-isei-08.htm](http://www.harryganzeboom.nl/isco08/qa-isei-08.htm), last accessed 17 March 2015.
- International Labour Organization (ILO) (2007), *Resolution concerning updating the International Standard Classification of Occupations*, [www.ilo.org/public/english/bureau/stat/isco/docs/resol08.pdf](http://www.ilo.org/public/english/bureau/stat/isco/docs/resol08.pdf), last accessed 17 March 2015.
- International Test Commission (2013), *ITC Guidelines on Test Use*, [https://www.intestcom.org/files/guideline\\_test\\_use.pdf](https://www.intestcom.org/files/guideline_test_use.pdf), last accessed 6 January 2017.
- Muraki, E. (1992), “A generalized partial credit model: Application of an EM algorithm”. *Applied Psychological Measurement*, No. 16, pp. 159-176.
- OECD (2017), *PISA 2015 Technical Report*, OECD, Paris, <http://www.oecd.org/pisa/data/2015-technical-report/>
- OECD (2014), *PISA 2012 Technical Report*, OECD, Paris, <http://www.oecd.org/pisa/pisa-products/PISA-2012-technical-report-final.pdf>.
- OECD (2012), *PISA 2009 Technical Report*, OECD, Paris, <http://www.oecd.org/pisa/pisa-products/50036771.pdf>.
- OECD (2009), *PISA 2006 Technical Report*, OECD, Paris, <http://www.oecd.org/pisa/pisa-products/42025182.pdf>.

- OECD (2009a), *PISA Data Analysis Manual. SAS®, Second Edition*, PISA, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264056251-en>.
- OECD (2009b), *PISA Data Analysis Manual. SPSS®, Second Edition*, PISA, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264056275-en>.
- von Davier, M. (2008), “A general diagnostic model applied to language testing data”, *British Journal of Mathematical and Statistical Psychology*, No 61, pp. 287-307.
- Wolter, K.M. (2007), *Introduction to Variance Estimation*, Second edition, Springer, New York, NY.
- Wu, M.L., Adams, R.J., & Wilson, M.R. (2007), *ACER ConQuest version 2.0: generalized item response modelling software*, Camberwell, VIC, Acer Press.





## *Chapter 4*

### School Report

#### 12. Supporting the preparation of a national version of the school report template

Schools that participate in the PISA-based Test for Schools receive a comprehensive report that are unique to each school and information on the learning climate at school, as well as students' engagement toward learning, to enhance school leaders' understanding of the overall performance of a school. The report also contains examples of relevant school policies and practices from around the world to stimulate reflections and discussions among local educators on school improvement. These reports are referred to as “eBooks”: to review a sample eBook, the reader can visit: [www.oecd.org/pisa/aboutpisa/PISA-based-Test-for-Schools-European-School-Culham.pdf](http://www.oecd.org/pisa/aboutpisa/PISA-based-Test-for-Schools-European-School-Culham.pdf).

The OECD is responsible for the preparation of a national version of the school report eBook template. All templates are prepared in English and follow the same general outline:

- Summary of your school's results
- Reader's guide
- Section 1 – Introduction: Understanding your school's results
- Section 2 – What students at your school know and can do in reading, mathematics and science
- Section 3 – The learning environment and student engagement at your School
- Section 4 – Your School's results in an international context
- Section 5 – Excellence and equity at your School
- References
- Annex A – The PISA-based Test for Schools in the [insert country name] [insert school year]
- Annex B – Overview of the PISA-based Test for Schools
- Annex C – Examples of test questions
- Annex D – Tables of results from PISA 2015 for countries and economies

### ***a. Proposal for adaptations***

Adaptations to the school report template can include, but are not limited to, changing the selection of countries/economies for comparison to your school/country results, selecting additional and/or different results to highlight in the report, modifying the figures, modifying the terminology used to describe types of schools (e.g., public and private schools may be referred to otherwise in other countries), etc. Adaptions to the report template are done by the OECD in consultation with and with the support of the National Service Provider.

### ***b. Translation of the template***

The National Service Provider is responsible for the translation of the school report template in the language(s) of use in the country. The following disclaimer is included on the legal page (page 4) of each school's report: "The translation of this OECD document was undertaken by [insert name of company/organisation that did the translation]. In the event of any discrepancy between the original work and the translation, only the text of original work shall be considered valid."

Once the template has been translated, it must be reviewed and approved by the OECD. The National Service Provider must allow at least one week for the review process. The OECD reviews the layout as well as data/figures and consults with the National Service Provider in the case of discrepancies between the English version and the translated version of the template. The OECD does not check the quality and validity of the translated text, as this is the responsibility of the National Service Provider.

## **13. Production of the school report eBooks**

The National Service Provider is responsible for producing each school report eBook based on the template that the OECD prepares in InDesign format. The OECD will also provide an Excel workbook containing templates of all figures to be reported in the eBook. Each sheet of the Excel workbook contains a template figure based on fixed input data. The National Service Provider will only have to enter the school's values as input data for the figures in the appropriate areas of the Excel sheets.

The following sections explain automation procedures as well as the process for receiving the OECD's approval for mass production of the eBooks.

### ***a. Automatizing the school report production***

The following elements can be automatized with a layout software in the report, which eliminates the need to manually enter information specific to each school's results and makes eBook production much more efficient:

- Cover pages (pages 1 and 3 of the eBook containing the name of the school)
- Headers of all pages of the report (containing the name of the school)



- Summary of your school’s results: text and figures
  - School name mentioned throughout summary
  - Means scores for reading, mathematics and science in Figure A and the text that follows
  - Language used to describe school performance compared to national performance (in the text after Figure A and in the section “Some highlights of results for your school”)
    - Statistically significantly higher/lower/is not significantly different from
    - Above the average; not significantly different from the average; below the average
    - More positive than; similar to; less positive than
    - Higher than; similar to; lower than
  - Percentages that appear in the text describing Figures B and C
- Figures
- Annex A

When setting up the automation procedures, attention must be paid to the variety of cases among schools and of extreme cases (large confidence interval going over the figure frame). Additionally, if a school has a very long name that might not fit on one line in the header on each page of the report, the name of the school should be abbreviated so that it fits within the template.

### ***b. Receiving OECD’s approval on the eBooks***

The OECD reviews the first 10 eBooks following the update of a school report. The 10 eBooks must be selected for review with the following criteria: schools must have different numbers of tested students (to see how the report automation accommodates for large confidence intervals), different levels of performance (to see how the report automation accommodates for results at the top and at the bottom of performance scales), different socioeconomic compositions (to see how the report automation accommodates for results at the top and at the bottom of the socioeconomic scale).

The National Service Provider should inform the OECD in advance of when they plan to send eBooks for review so that the OECD can schedule time for the review. The National Service Provider must allow a week for the OECD to review 10 eBooks. If the OECD receives the eBooks without having been informed in advance of when they will be available for review, the National Service Provider will have to allow more than one week before hearing back from the OECD.

Once the first 10 eBooks are reviewed, the National Service Provider can proceed with the production of the remaining eBooks. It is the National Service Provider's responsibility to distribute the eBooks to each school.

#### **14. Grouped school report**

The National Service Provider may also develop a grouped school report based on districts, regions or countries. The purpose of the grouped school report will never be to overlap with or replace the main PISA assessment, and the focus of the assessment will always be the schools. The grouped school report, as well as any other additional request to the standard School Report presented above, needs to be discussed and agreed upon with the OECD, based on the availability of the data and how robust the results are. The grouped school reports may have special sample requirements that are not specified in this Technical Report, as well as some analytical adaptations.

## Annex 1

### Technical documentation

List of technical documentation to be supplied by the OECD to the National Service Provider:

- School Coordinator Manual in English
- Test Administrator Manual in English
- Template of student tracking form
- Template of session attendance form
- Source versions of the 7 test booklets
- Source Student Questionnaire
- Source School Questionnaire
- Note-version of the Student Questionnaire
- Note-version of the School Questionnaire
- Coding guides for the cognitive items
- PFS Translation and Adaptation Guidelines
- PISA Student Questionnaire Codebook
- PISA School Questionnaire Codebook
- PFS data template and codebooks
- PFS international item parameters for cognitive items
- SAS® and SPSS® programmes for mapping ISCO-08 Occupation code to ISEI-08
- Input and output files of the latent class analysis performed with Mplus software on PISA 2009 data to generate student reader profiles
- ACER PFS Technical Report
- Template of InDesign files for the eBook production

Templates of Excel figures displaying schools' results





## Annex 2

### Examples of R codes for scaling

#### Example to scale reading items

```
library(TAM)

setwd("C:\\ PBTS \\ReadingIRT")
getwd()

Data1<-read.csv("ReadingIRT.csv ", header=T, sep=";", dec=",")

IDs = Data1[,1] # first 1 col are Ids
reading = Data1[,2:47] # the 46 reading items are in cols 2-47
reading[reading==6] <- NA # 9s already set to zero, 6s are not administered so set to missing
reg_READ = Data1[,c(142:146, 148:445)] # all the remaining variables are in the regression part of the model.

anchorValues = read.table("ReadingParameters.csv",header = T, sep=";",dec = ",")
model <- tam(reading, xsi.fixed = anchorValues, Y = reg_READ, pid=NULL)

#PVs
readingPVs = tam.pv(model_R, nplausible = 5)
readingPVs$pv

readingResults= cbind(IDs,(((0.883*(readingPVs$pv[2:6])-0.4837)/1.1002)*100+500))
readingResults<-cbind(readingResults,data.frame(Means=rowMeans(readingResults[,-1])))
dim(readingResults)

names(readingResults)<
c("STUDID","PV1_Read","PV2_Read","PV3_Read","PV4_Read","PV5_Read","Mean_Read")

write.table(readingResults, file="READPV.csv",sep = ";",dec=",", row.names = FALSE,
col.names = TRUE)
```

**Example to compute Mathematics Self-Efficacy (MATHEFF)**

```

library(TAM)

setwd("C:\\ PBTs \\MatheffIRT ")
getwd()

indices=read.csv("PISA12_STQUES.csv", header=T, sep=";", dec=",")
names(indices)

#first 1 col are IDs
IDs = indices[,5]

#cols 41-48 are Qs for MATHEFF
i_matheff = indices[,81:88]
names(i_matheff)

# MATHEFF reverse coded
MATHEFFrecoded = i_matheff
MATHEFFrecoded[i_matheff==1] <- 3 #recode (1=3)(2=2)(3=1)(4=0)
MATHEFFrecoded[i_matheff==2] <- 2
MATHEFFrecoded[i_matheff==3] <- 1
MATHEFFrecoded[i_matheff==4] <- 0
MATHEFFrecoded[i_matheff==7] <- NA
MATHEFFrecoded[i_matheff==8] <- NA
MATHEFFrecoded[i_matheff==9] <- NA

# Create a vector with the anchor values
anchorValues = rbind(-2.02076,-0.26592,1.79057,
  -2.01532,-0.37606,1.44041,
  -1.70363,0.28183,1.82278,
  -1.94462,-0.27570,1.75517,
  -1.71392,-0.67904,0.52663,
  -1.34047,0.84621,2.16226,
  -1.46061,-0.01307,1.24148,
  -1.32062,0.78500,2.47740)

#Combine into a matrix
anchorValues <- cbind( 1:(length(anchorValues)) , anchorValues)

# IRT model
MATHEFFmodel <- tam(MATHEFFrecoded, xsi.fixed=anchorValues)
MATHEFFwle <- tam.wle(MATHEFFmodel)

MATHEFFmodel$xsi
MATHEFFmodel$item

```

```
MATHEFFmodel$person
MATHEFFmodel$item

plot(MATHEFFmodel)

# item response curves
plot(MATHEFFmodel, items=1:5, type="items" , export=FALSE)

# Standardise values
EndResults = cbind(IDs,(MATHEFFwle$theta-1.15)/1.5)

write.table(EndResults,file="MATHEFF_12.csv",sep = ";",dec="," ,row.names = FALSE,
col.names = TRUE)
```



## **Organisation for Economic Co-operation and Development**

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalization. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an aging population. The organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Commission takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.



# TECHNICAL REPORT

The OECD PISA-Based for Schools is an assessment designed for use by schools and groups of schools around the world. It aims to support international benchmarking, school improvement efforts, and research. Based on the triennial PISA studies, the assessment collects information about 15-year old students' applied knowledge and skills in reading, mathematics and science as well as their attitudes towards learning and school. The data collected by the assessment is a valuable source of information for school principals, educators, parents and students, helping them to make evidence-based decisions regarding educational provision.

This Technical Report describes the methodologies and procedures used to implement the paper-based version of the PISA-based Test for Schools. The purpose of the report is two-fold: first the report serves as documentation of how the assessment is implemented during 2016-2019, and second it serves as a reference manual for the continued implementation of the assessment across the world.

For further information please visit our website:  
[www.oecd.org/pisa/aboutpisa/pisa-based-test-for-schools.htm](http://www.oecd.org/pisa/aboutpisa/pisa-based-test-for-schools.htm)

or contact us at:  
[pisabasedtestforschools@oecd.org](mailto:pisabasedtestforschools@oecd.org)