



12

Scaling outcomes

Results of the IRT scaling and population modeling	226
Transforming the plausible values to PISA scales.....	233
Linking error	237
International characteristics of the item pool	237

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.



This chapter illustrates the outcomes of applying the item response theory (IRT) scaling and population model for the generation of plausible values to the PISA 2015 main survey assessment data. In the item response theory (IRT) scaling stage, all available items and data from prior PISA cycles (2006, 2009, 2012) were scaled together with the 2015 data via a concurrent calibration using country-by-language-by-cycle groups. However, only results based on the item parameters for the 2015 items are presented here.

RESULTS OF THE IRT SCALING AND POPULATION MODELING

The linking design for the PISA main survey was aimed at establishing comparability across countries, languages, assessment modes (paper-based and computer-based assessments), and between the 2015 PISA cycle and previous PISA cycles (as far back as 2006, which had been the last time that science was the major domain). By imposing constraints on the item parameters in the item response scaling, the estimated parameters for trend and new items were placed on the same scale, along with items that were used in previous PISA cycles (but not selected for 2015). An additional outcome of the item response theory scaling is that paper-based (PBA) and computer-based (CBA) assessment items can be placed on the same scale. The items generally fit well across countries, allowing for the use of common international item parameters. These international (or common) parameters are what allow for comparability of results across countries and years. However, there are cases where the international item parameters for a given item do not fit well for a particular country or language group, or subset of countries or language groups. In these instances (i.e. when there is item misfit), which imply interactions in certain groups (e.g. item-by-country/language interactions, item-by-mode interactions, item-by-cycle interactions), item constraints were released to allow the estimation of unique item parameters. This was done for a relatively small number of cases across items and groups.

Unique item parameter estimation and national item deletion

The item response theory calibration for the PISA 2015 main survey data was carried out separately for each of the PISA 2015 domains (reading, mathematical, science, financial literacy, and collaborative problem solving). Both science (as the main domain in PISA 2015) and collaborative problem solving (CPS) (as a new domain in PISA 2015) included new items; science also included trend items. All of the other domains included trend items only. Item fit was evaluated using the mean deviation and the root mean squared deviation. Both deviations were calculated for all items in each country-language group for each mode and PISA cycle.

The final item parameters were estimated based on a concurrent calibration using the data from PISA 2015 as well as from previous PISA cycles going back to 2006. There were only a few items in mathematics and collaborative problem solving that had to be excluded from the item response theory analyses (in all country-by-language-by-cycle groups) due to either almost no response variance, scoring or technical issues (either problems with the delivery platform or with the coding on the platform), or very low or even negative item total correlations; Table 12.1 gives an overview of these items.

Table 12.1 Items that were excluded from the IRT analyses

Domain	Item	Mode	Reason
Maths (1 item)	CM192Q01	CBA	Technical issue
CPS (4 items)	CC104104	CBA	Very few responses in category 0
	CC104303		Technical issue
	CC102208		Very few responses in category 0
	CC105405		Low and negative item-total correlation (correlation close to zero)

Note: The problems observed for the items in the table were shown over all countries.

The international/common item parameters and unique national item parameters were estimated for each domain using unidimensional multigroup item response theory models. For analysis purposes, the international/common item parameters are divided into two groups: scalar invariant and metric invariant parameters. Scalar invariant items correspond to items where the slope and threshold parameters are constrained to be the same in both paper-based and computer-based modes. Metric invariant items correspond to items where the slope is constrained to be the same, but the threshold differs across modes. For new items from science and collaborative problem solving, there are no metric invariant item parameters because these were administered only as part of the computer-based assessment; for financial literacy, all items were constrained to be scalar invariant. As such, only scalar invariant percentages are reported in these domains. For each domain, the scalar and metric invariant item parameters represent the stable linked items between the previous and PISA 2015 scales; the unique parameters are included to reduce measurement error. Table 12.2 shows



the percentage of common and unique item parameters by domain computed by dividing the number of unique item-by-country cells through the total item-by-country cells. Note that the percentage of scalar/metric invariant international/common item parameters was above 90% in cognitive domains with the exception of reading and science. Further, only a small number of items received unique item parameters (either group-specific or the same parameters across a subset of groups) except for reading. In reading, the proportion of scalar/metric invariant international/common item parameters was 89.01%, the proportion of group-specific item parameters was 3.01%, and 7.98% received the same unique item parameters across a subset of countries. For trend items in science, 89.70% received scalar/metric invariant international/common item parameters, while 2.62% received group-specific item parameters, and 7.68% received the same parameters across a subset of countries.

Table 12.2 Percentage of common and unique item parameters in each domain for PISA 2015

	Maths	Reading	Science trend	Science new	CPS	Financial literacy
% of unique item parameters (group-specific)	2.16%	3.01%	2.62%	2.05%	1.85%	4.40%
% of unique item parameters (same parameters across a subset of groups)	3.36%	7.98%	7.68%	4.60%	3.19%	2.69%
% of metric invariant common/international item parameters	33.22%	30.33%	20.96%	N/A	N/A	N/A
% of scalar invariant common/international item parameters	61.25%	58.68%	68.74%	93.35%	94.96%	92.91%
Mode and number of items in the PISA 2015 main survey	PBA: 83 items, CBA: 81 items	PBA: 103 items, CBA: 103 items	PBA: 85 items, CBA: 85 items	CBA: 99 items	CBA: 117 items	CBA: 43 items

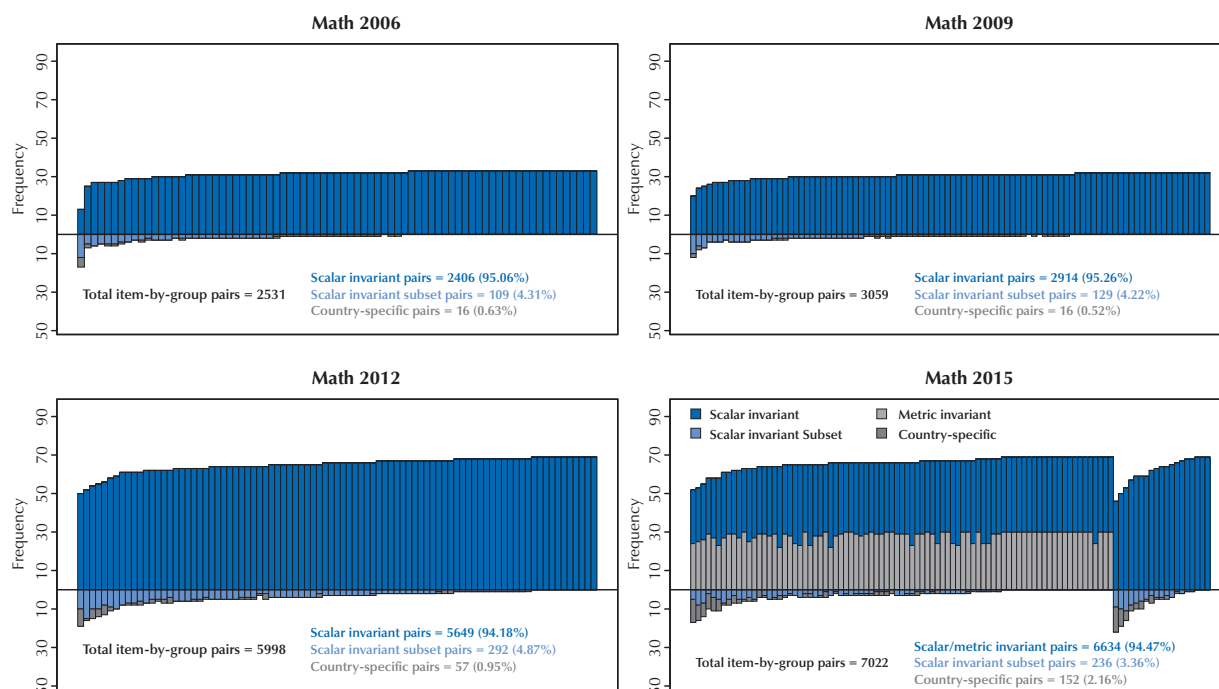
Note: Interactions go across modes and cycles; Kazakhstan is not included due to adjudication issues.

An overview of the proportions of international/common (invariant) item parameters and group-specific item parameters in each domain for each relevant assessment cycle is given in Figures 12.1 to 12.6. The figures also provide an overview of the proportion of scalar invariant item parameters (items sharing common difficulty and slope parameters across modes) and partially or metric invariant item parameters (items sharing common slope parameters across modes) with regard to the mode effect modeling described in Chapter 9: dark blue indicates scalar invariant item parameters, light grey (the lighter grey above the horizontal line) indicates metric invariant item parameters, medium blue indicates scalar invariant item parameters for a subset of groups (unique parameters different from the common parameter, but for several groups sharing the same unique parameter), and dark grey indicates group-specific item parameters. In addition, Annex H provides information about which trend items are scalar invariant and which are partially or metric invariant for each cognitive domain. Recall that both scalar and metric invariant item parameters (dark blue and light grey) contribute to improve the comparability across groups, while unique item parameters (medium blue and dark grey) contribute to the reduction of measurement error. Across every cycle and every domain, it is clear that international/common (invariant) item parameters dominate and only a small proportion of the item parameters are group-specific (i.e. dark grey). Results show that the overall item fit in each domain for each group is very good, resulting in a small numbers of unique item parameters and high comparability of the data. There was no consistent pattern of deviations for any one particular country-by-language group. The results also illustrate that the trend items show good fit, ensuring the quality of the trend measure across different assessment cycles (2015 data versus 2006-2012), different assessment modes (PBA versus CBA), and even across different countries and languages. An overview of the number of deviations per item across all country-by-language-by-cycle groups for items in each domain is given in Annex G.

After the IRT scaling was finalised, item parameter estimates were delivered to each country, including an indication of which items received international/common item parameters and which received unique item parameters. Table 12.3 gives an example of the information provided to countries: the first column shows the domain; the second column shows the flag that indicates whether an item received a unique parameter or was excluded from the IRT scaling; and the remaining columns show the final item parameter estimates (for each item, the slope, difficulty and threshold parameters for polytomous items were listed). A slope parameter of 1 indicates that a Rasch model was fitted for these items; slope estimates different from 1 indicate that the two-parameter logistic model (2PLM) was fitted.

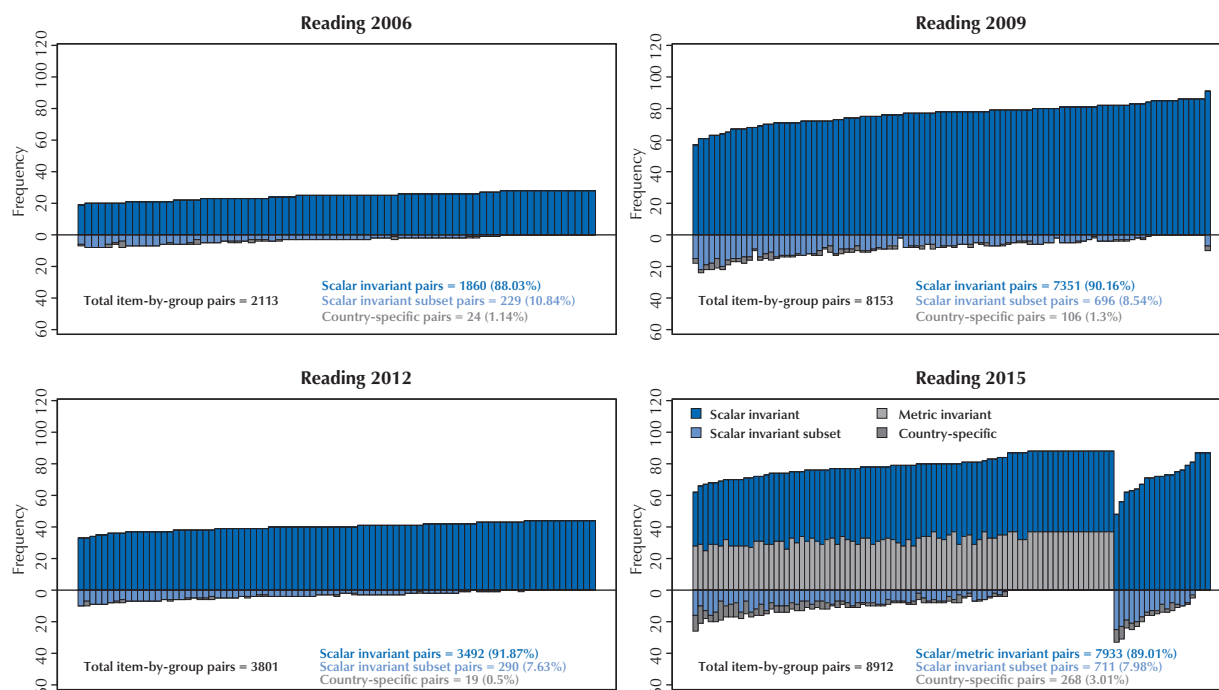
■ Figure 12.1 ■

Frequencies of international (invariant) and unique item parameters in maths (note that frequencies were counted using item-by-group pairs)



■ Figure 12.2 ■

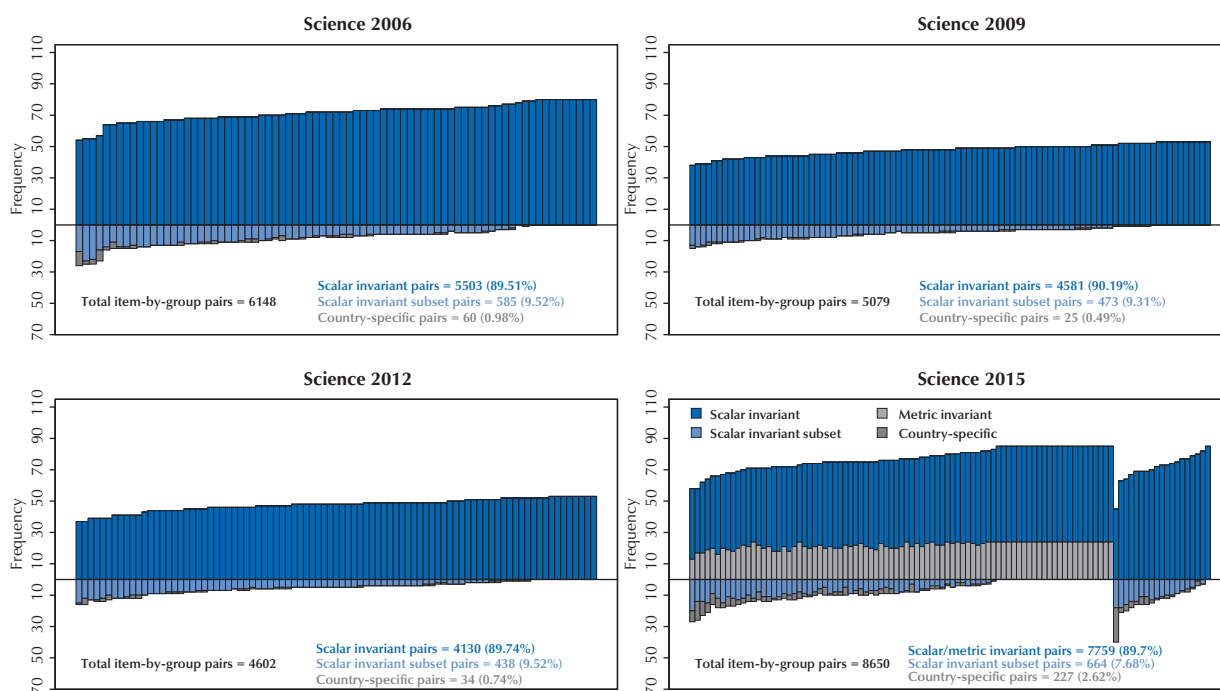
Frequencies of international (invariant) and unique item parameters in reading (note that frequencies were counted using item-by-group pairs)





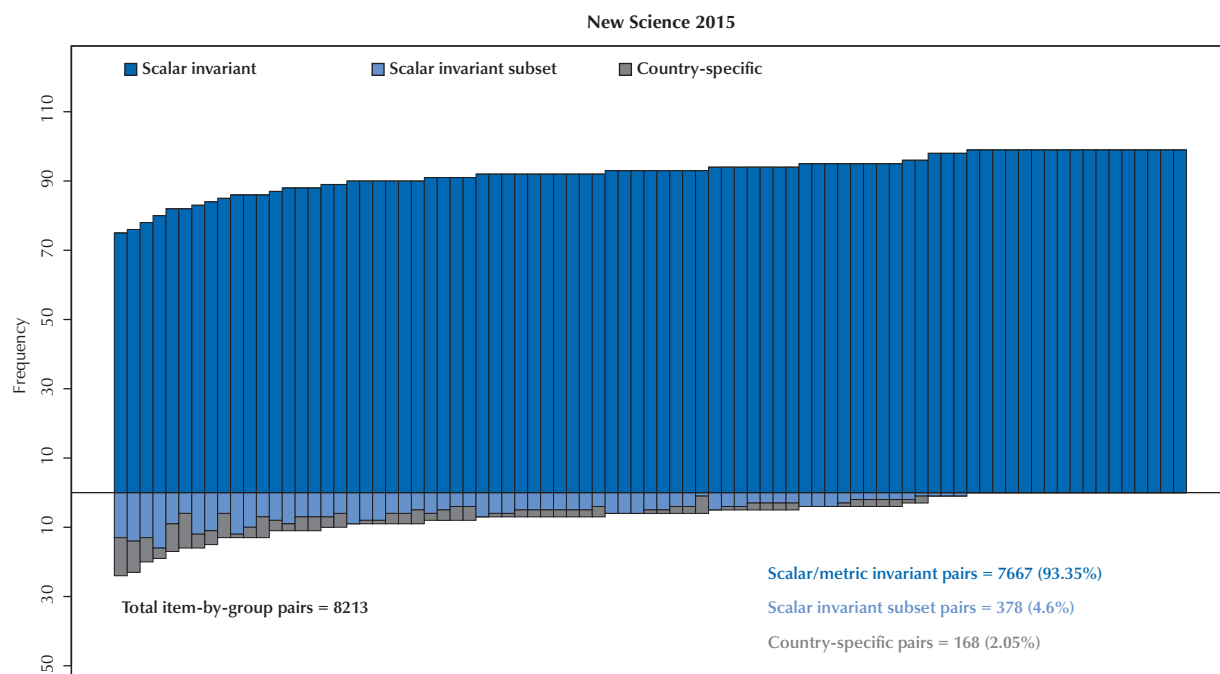
■ Figure 12.3 ■

Frequencies of international (invariant) and unique item in trend science (note that frequencies were counted using item-by-group pairs)



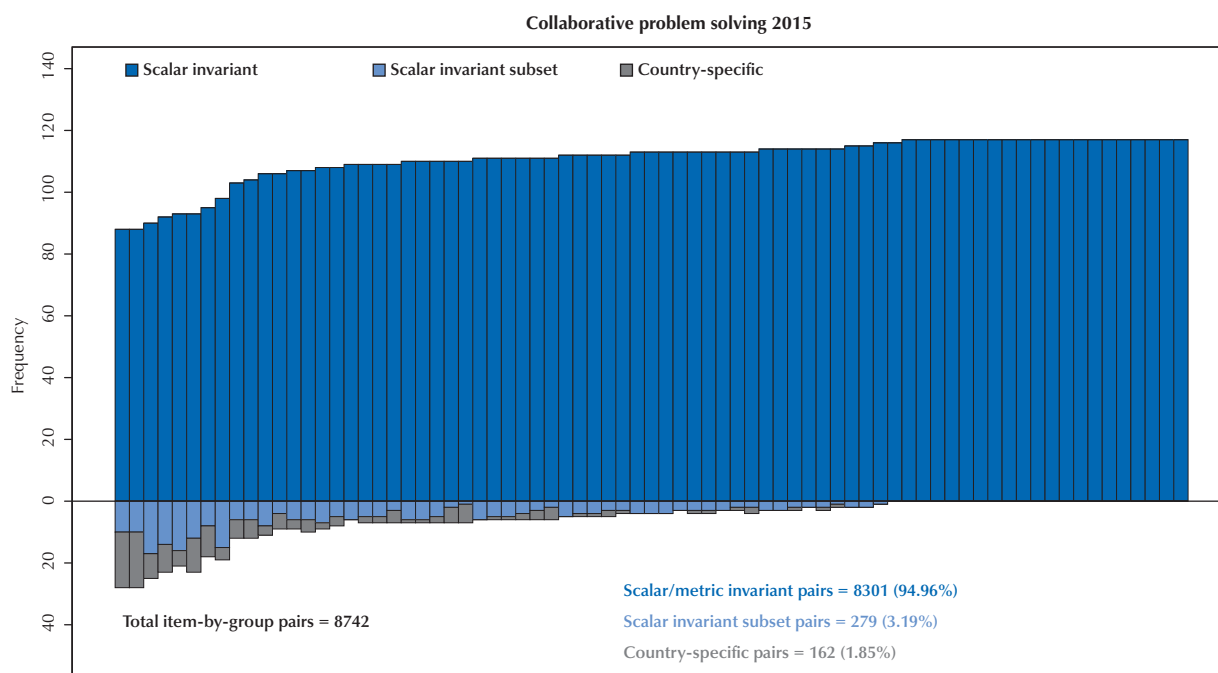
■ Figure 12.4 ■

Frequencies of international (invariant) and unique item in new science (note that frequencies were counted using item-by-group pairs)



■ Figure 12.5 ■

Frequencies of international (invariant) and unique item in CPS (note that frequencies were counted using item-by-group pairs)



■ Figure 12.6 ■

Frequencies of international (invariant) and unique item in financial literacy (note that frequencies were counted using item-by-group pairs)

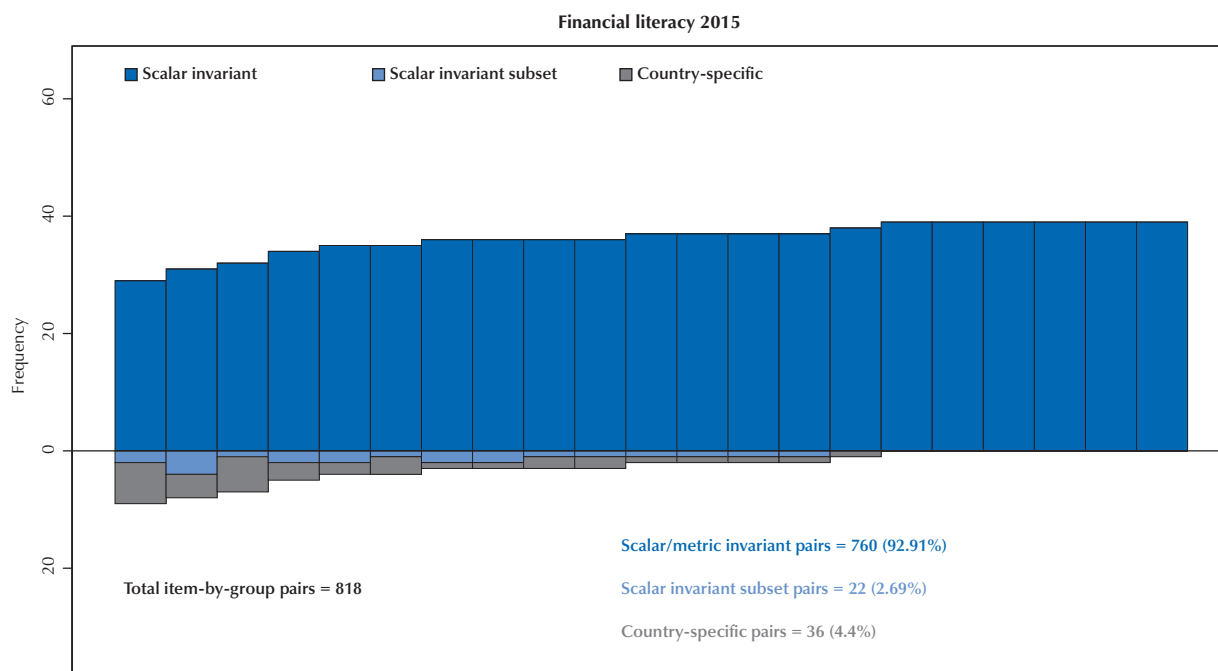




Table 12.3 Example of table for item parameter estimates provided to the countries

Domain	Flag	Item	Slope	Difficulty	IRT_Step1	IRT_Step2
Maths		PM00GQ01	1	1.62226		
Maths		PM00KQ02	1	1.11572		
Maths		PM033Q01S	1	-0.95604		
Maths		PM034Q01S	1	0.15781		
Maths	Unique item parameters	PM155Q01	1.42972	-0.35538		
Maths		PM155Q02	1	-0.35727	-0.42436	0.42436
Maths		PM155Q03	1.08678	0.73497	-0.20119	0.20119
Maths		PM155Q04S	1	-0.27556		
Maths		PM192Q01S	1	0.20948		
Maths	Excluded from scaling	PM936Q01				

Generating student scale scores and reliability of the PISA scales

Given the rotated and incomplete assessment design, it is not possible to calculate marginal reliabilities for each cognitive domain. In order to get an indication of test reliability, the explained variance (i.e. variance explained by the model) for each cognitive domain was computed based on the weighted posterior variance. The variance is computed using all 10 plausible values as follows: $1 - (\text{expected error variance} / \text{total variance})$. The weighted posterior variance is an expression of the posterior measurement error and is obtained through the population modeling. The expected error variance is the weighted average of the posteriori variance. This term was estimated using the weighted average of the variance of the plausible values (the posteriori variance is the variance across the 10 plausible values). The total variance was estimated using a resampling approach (Efron, 1982). It was estimated for each country depending on the country-specific proficiency distributions for each cognitive domain.

Applying the conditioning approach described in Chapter 9 and anchoring all of the item parameters at the values obtained from the final IRT scaling, plausible values were generated for all sampled students. Table 12.4 gives the median of national reliabilities for the generated scale scores based on all 10 plausible values. National reliabilities of the main cognitive domains based on all 10 plausible values are presented in Table 12.5.

Table 12.4 Reliabilities of the PISA cognitive domains and Science subscales overall countries¹

Mode	Domains	Median	S.D.	Max	Min
CBA	Maths	0.85	0.03	0.90	0.75
	Reading	0.87	0.02	0.90	0.80
	Science	0.91	0.02	0.93	0.82
	CPS	0.78	0.03	0.83	0.70
	Financial literacy	0.83	0.06	0.93	0.72
	Science subscales				
	Explain phenomena scientifically	0.89	0.03	0.91	0.80
	Evaluate and design scientific inquiry	0.87	0.04	0.90	0.71
	Interpret data and evidence scientifically	0.89	0.03	0.92	0.78
	Content	0.89	0.02	0.91	0.81
	Procedural & epistemic	0.90	0.03	0.92	0.78
	Earth & science	0.88	0.03	0.90	0.77
	Living	0.89	0.03	0.91	0.79
	Physical	0.88	0.03	0.91	0.76
PBA	Maths	0.80	0.05	0.87	0.67
	Reading	0.82	0.04	0.88	0.72
	Science	0.86	0.04	0.92	0.77

1. Please note that Argentina, Malaysia, and Kazakhstan were not included in this analysis due to adjudication issues (inadequate coverage of either population or construct).

[Part 1/2]

Table 12.5 National reliabilities for main cognitive domains

Mode	Country/economy	Maths	Reading	Science	CPS	Financial literacy
CBA	Australia	0.84	0.86	0.92	0.76	0.93
CBA	Austria	0.87	0.88	0.93	0.80	–
CBA	Belgium	0.89	0.89	0.93	0.80	0.87
CBA	Brazil	0.78	0.82	0.87	0.71	0.72
CBA	Bulgaria	0.85	0.88	0.92	0.82	–
CBA	Canada	0.83	0.83	0.91	0.74	0.76
CBA	Chile	0.86	0.86	0.91	0.78	0.83
CBA	B-S-J-G (China) ¹	0.90	0.90	0.93	0.83	0.88
CBA	Colombia	0.82	0.87	0.89	0.76	–
CBA	Costa Rica	0.78	0.83	0.85	0.70	–
CBA	Croatia	0.86	0.87	0.91	0.76	–
CBA	Cyprus ²	0.84	0.84	0.90	0.74	–
CBA	Czech Republic	0.88	0.88	0.92	0.77	–
CBA	Denmark	0.85	0.86	0.91	0.78	–
CBA	Dominican Republic	0.81	0.86	0.84	–	–
CBA	Estonia	0.85	0.86	0.91	0.79	–
CBA	Finland	0.85	0.87	0.91	0.77	–
CBA	France	0.88	0.89	0.93	0.77	–
CBA	Germany	0.86	0.86	0.92	0.76	–
CBA	Greece	0.86	0.87	0.91	0.79	–
CBA	Hong Kong (China)	0.84	0.85	0.90	0.77	–
CBA	Hungary	0.88	0.89	0.92	0.81	–
CBA	Iceland	0.83	0.86	0.91	0.76	–
CBA	Ireland	0.85	0.87	0.91	–	–
CBA	Israel	0.87	0.88	0.92	0.83	–
CBA	Italy	0.87	0.87	0.91	0.80	0.81
CBA	Japan	0.85	0.85	0.91	0.75	–
CBA	Korea	0.85	0.85	0.91	0.78	–
CBA	Latvia	0.85	0.86	0.90	0.75	–
CBA	Lithuania	0.85	0.87	0.91	0.80	0.83
CBA	Luxembourg	0.87	0.89	0.93	0.77	–
CBA	Macao (China)	0.82	0.86	0.90	0.78	–
CBA	Malaysia	0.86	0.87	0.90	0.79	–
CBA	Mexico	0.79	0.84	0.86	0.75	–
CBA	Montenegro	0.80	0.84	0.88	0.74	–
CBA	Netherlands	0.89	0.89	0.93	0.79	0.88
CBA	New Zealand	0.85	0.86	0.92	0.79	–
CBA	Norway	0.84	0.85	0.91	0.75	–
CBA	Peru	0.82	0.88	0.87	0.78	0.87
CBA	Poland	0.86	0.87	0.92	–	0.83
CBA	Portugal	0.87	0.86	0.92	0.78	–
CBA	Qatar	0.85	0.89	0.91	–	–
CBA	Russia	0.78	0.80	0.88	0.75	0.73
CBA	Singapore	0.87	0.88	0.93	0.79	–
CBA	Slovak Republic	0.86	0.89	0.92	0.77	0.76
CBA	Slovenia	0.88	0.89	0.93	0.79	–
CBA	Spain	0.86	0.86	0.91	0.75	0.81
CBA	Sweden	0.85	0.86	0.92	0.78	–
CBA	Switzerland	0.86	0.88	0.92	–	–
CBA	Chinese Taipei	0.87	0.88	0.93	0.78	–
CBA	Thailand	0.81	0.86	0.88	0.83	–
CBA	Tunisia	0.75	0.80	0.82	0.70	–
CBA	Turkey	0.82	0.85	0.89	0.74	–
CBA	United Arab Emirates	0.83	0.87	0.91	0.80	–
CBA	United Kingdom	0.87	0.88	0.92	0.83	–
CBA	United States	0.87	0.88	0.92	0.81	0.87
CBA	Uruguay	0.85	0.87	0.90	0.78	–



[Part 2/2]
Table 12.5 National reliabilities for main cognitive domains

Mode	Country/economy	Maths	Reading	Science	CPS	Financial literacy
PBA	Albania	0.75	0.79	0.84	–	–
PBA	Algeria	0.67	0.72	0.77	–	–
PBA	Argentina	0.79	0.82	0.85	–	–
PBA	FYROM	0.79	0.79	0.84	–	–
PBA	Georgia	0.83	0.83	0.86	–	–
PBA	Indonesia	0.78	0.77	0.82	–	–
PBA	Jordan	0.78	0.82	0.86	–	–
PBA	Kazakhstan	0.73	0.71	0.78	–	–
PBA	Kosovo	0.80	0.81	0.82	–	–
PBA	Lebanon	0.82	0.85	0.86	–	–
PBA	Malta	0.87	0.88	0.92	–	–
PBA	Moldova	0.78	0.83	0.86	–	–
PBA	Romania	0.80	0.82	0.86	–	–
PBA	Trinidad and Tobago	0.86	0.84	0.88	–	–
PBA	Viet Nam	0.83	0.84	0.87	–	–

1. B-S-J-G (China) data represent the regions of Beijing, Shanghai, Jiangsu, and Guangdong.

2. 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 recognizes 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.

The table above shows that the explained variance by the combined IRT and latent regression model (population or conditioning model) is at a comparable level across countries. While the population model reaches levels of above 0.80 for reading, mathematics and science, it is important to keep in mind that this is not to be confused with a classical reliability coefficient, as it is based on more than the item responses. Comparisons among individual students are not appropriate because the apparent accuracy of the measures is obtained by statistically adjusting the estimates based on background data. This approach does provide improved behavior of subgroup estimates, even if the plausible values obtained using this methodology are not suitable for comparisons of individuals (e.g. Mislevy & Sheehan, 1987; von Davier et al., 2006).

TRANSFORMING THE PLAUSIBLE VALUES TO PISA SCALES

The plausible values were transformed using a linear transformation to form a scale that is linked to the historic PISA scale. This scale can be used to compare the overall performance of countries or subgroups within a country.

For science, reading and mathematics, country results from the 2006, 2009 and 2012 PISA cycles for OECD countries were used to compute the transformation coefficients for each content domain separately. The country means and variances used to compute the transformation coefficients included only those values from the cycle in which a given content domain was the major domain. Hence, the transformation coefficients for science are based on the 2006 reported and model-based results, reading coefficients are based on the 2009 results, and mathematics coefficients are based on the 2012 results. Only the results for countries designated as OECD countries in the respective PISA reporting cycle were used to compute the transformation coefficients. If m_{vij} is the reported mean for country i in cycle j , m_{xij} is the model-based mean obtained from the concurrent calibration using the software *mdltn*, and s_{vij}^2 and s_{xij}^2 are the reported and model-based score variances respectively. The same transformation was used for all plausible values (within a given domain). The transformation coefficients for a given content domain were computed as:

12.1

$$A = \frac{\tau_{vj}}{\tau_{xj}}$$

12.2

$$B = \bar{m}_{vj} - A\bar{m}_{xj}$$

12.3

$$\tau_{Yj} = \sqrt{\tau_{Yj}^2} = \sqrt{\left[\frac{1}{n_j - 1} \sum_{i=1}^{n_j} (m_{Yij} - \bar{m}_{Yj})^2 \right] + \frac{1}{n_j} \sum_{i=1}^{n_j} S_{Yij}^2}$$

12.4

$$\tau_{Xj} = \sqrt{\tau_{Xj}^2} = \sqrt{\left[\frac{1}{n_j - 1} \sum_{i=1}^{n_j} (m_{Xij} - \bar{m}_{Xj})^2 \right] + \frac{1}{n_j} \sum_{i=1}^{n_j} S_{Xij}^2}$$

where = $\begin{cases} 2006 & \text{Science} \\ 2009 & \text{Reading} \\ 2012 & \text{Maths} \end{cases}$

The values \bar{m}_{Yj} and \bar{m}_{Xj} are grand means of the reported and model-based country means in cycle j , respectively. The terms τ_{Yj}^2 and τ_{Xj}^2 correspond to the total variance, defined as the variance of the country means, plus the mean of the country variances respectively. The square root of these terms is taken to compute the standard deviations τ_{Yj} and τ_{Xj} . The 2015 plausible values (PVs) for examinee k in country i were transformed to the PISA scale via the following transformation:

12.5

$$PV_{Tik} = A \times PV_{Uik} + B$$

The subscripts T and U correspond to the transformed and untransformed values respectively.

For financial literacy, country results from the 2012 PISA cycle were used to compute the transformation coefficients. The method used to compute the coefficients is the same as that used for reading, mathematics and science. The key distinction is that in reading, mathematics and science, only results for OECD countries were used to compute the coefficients, whereas, for financial literacy, all available country data were used to compute the coefficients. This decision was made because there were too few OECD countries to provide a defensible transformation of the results. The plausible values for financial literacy were transformed using the same linear transformation as for reading, mathematics and science.

A new scale for CPS was established in PISA 2015. Consistent with the introduction of content domains in previous PISA cycles, transformation coefficients for CPS were computed such that the plausible values for OECD countries have a mean of 500 and a standard deviation of 100. The 10 sets of plausible values were stacked together and the weighted mean and variance (and by extension SD) were computed. Stated differently, the full set of transformed plausible values for CPS have a weighted mean of 500 and a weighted SD of 100 (based on senate weights).

If X_{kv} is the v^{th} PV $\{v \text{ in } 1, 2, \dots, 10\}$ for examinee k , the transformation coefficients for CPS are computed as

12.6

$$A = \frac{100}{\tau_{PV}}$$

12.7

$$B = 500 - A[\bar{X}_{kv}] = 500 - A \left[\frac{\sum_{v=1}^{10} \sum_{k=1}^n X_{kv} W_{kv}}{10 \sum_{k=1}^n W_{kv}} \right]$$

12.8

$$\tau_{PV} = \sqrt{\tau_{PV}^2} = \sqrt{\frac{\sum_{v=1}^{10} \sum_{k=1}^n W_{kv} (X_{kv} - \bar{X}_{kv})^2}{[(10n - 1) \sum_{k=1}^n W_{kv}] / n}}$$



The grand mean of the PVs, \bar{X}_{kv} , was computed by compiling all 10 sets of PVs into a single vector (the corresponding senate weights were compiled in a separate vector) then finding the weighted mean of these values. The weighted variance, τ_{PV}^2 , was computed using the vector of PVs as well. The square root is taken to compute the standard deviation, τ_{PV} . The plausible values for CPS were transformed using the same approach as that for science, reading, mathematics and financial literacy. The transformations for reading, mathematics, science and financial literacy used the model-based results from the concurrent calibration (IRT scaling) in order to align the results with previously established scales. The transformation for CPS is based on the PVs because this is the first time the results for this domain have been scaled.

The transformation coefficients for all content domains are presented in Table 12.6. The A coefficient adjusts the variability (standard deviation) of the resulting scale while the B coefficient adjusts the scale location (mean).

Table 12.6 PISA 2015 transformation coefficients

Domain	A	B
Science	168.3189	494.5360
Reading	131.5806	437.9583
Mathematics	135.9030	514.1848
Financial literacy	140.0807	490.7259
Collaborative problem solving	196.7695	462.8102

Table 12.7 shows the average transformed plausible values for each cognitive domain by country as well as the resampling-based standard errors.

[Part 1/2]

Average plausible values (PVs) and resampling-based standard errors (SE) by country/economy for the PISA domains of science, reading, mathematics, financial literacy, and collaborative problem solving (CPS)

Table 12.7

Country/economy	Maths		Reading		Science		CPS		Financial literacy	
	Average PV	SE	Average PV	SE	Average PV	SE	Average PV	SE	Average PV	SE
International average	462	0.32	461	0.34	466	0.31	486	0.36	481	0.95
Albania	413	3.45	405	4.13	427	3.28				
Algeria	360	2.95	350	3.00	376	2.64				
Argentina	409	3.05	425	3.22	432	2.87				
Australia	494	1.61	503	1.69	510	1.54	531	1.91	504	1.91
Austria	497	2.86	485	2.84	495	2.44	509	2.56		
Belgium	507	2.35	499	2.42	502	2.29	501	2.39	541	2.95
Brazil	377	2.86	407	2.75	401	2.30	412	2.30	393	3.84
B-S-J-G (China)	531	4.89	494	5.13	518	4.64	496	3.97	566	6.04
Bulgaria	441	3.95	432	5.00	446	4.35	444	3.85		
Canada	516	2.31	527	2.30	528	2.08	535	2.27	533	4.62
Chile	423	2.54	459	2.58	447	2.38	457	2.69	432	3.74
Colombia	390	2.29	425	2.94	416	2.36	429	2.30		
Costa Rica	400	2.47	427	2.63	420	2.07	441	2.42		
Croatia	464	2.77	487	2.68	475	2.45	473	2.52		
Cyprus ¹	437	1.72	443	1.66	433	1.38	444	1.71		
Czech Republic	492	2.40	487	2.60	493	2.27	499	2.20		
Denmark	511	2.17	500	2.54	502	2.38	520	2.53		
Dominican Republic	328	2.69	358	3.05	332	2.58				
Estonia	520	2.04	519	2.22	534	2.09	535	2.47		
Finland	511	2.31	526	2.55	531	2.39	534	2.55		
France	493	2.10	499	2.51	495	2.06	494	2.42		
FYROM	371	1.28	352	1.41	384	1.25				
Georgia	404	2.78	401	2.96	411	2.42				
Germany	506	2.89	509	3.02	509	2.70	525	2.85		
Greece	454	3.75	467	4.34	455	3.92	459	3.60		
Hong Kong (China)	548	2.98	527	2.69	523	2.55	541	2.95		
Hungary	477	2.53	470	2.66	477	2.42	472	2.35		
Iceland	488	1.99	482	1.98	473	1.68	499	2.26		

[Part 2/2]

Average plausible values (PVs) and resampling-based standard errors (SE) by country/economy for the PISA domains of science, reading, mathematics, financial literacy, and collaborative problem solving (CPS)

Table 12.7

Country/economy	Maths		Reading		Science		CPS		Financial literacy	
	Average PV	SE	Average PV	SE	Average PV	SE	Average PV	SE	Average PV	SE
Indonesia	386	3.08	397	2.87	403	2.57				
Ireland	504	2.05	521	2.47	503	2.39				
Israel	470	3.63	479	3.78	467	3.44	469	3.62		
Italy	490	2.85	485	2.68	481	2.52	478	2.53	483	2.80
Japan	532	3.00	516	3.20	538	2.97	552	2.68		
Jordan	380	2.65	408	2.93	409	2.67				
Kazakhstan	460	4.28	427	3.42	456	3.67				
Korea	524	3.71	517	3.50	516	3.13	538	2.53		
Kosovo	362	1.63	347	1.57	378	1.70				
Latvia	482	1.87	488	1.80	490	1.56	485	2.26		
Lebanon	396	3.69	347	4.41	386	3.40				
Lithuania	478	2.33	472	2.74	475	2.65	467	2.46	449	3.15
Luxembourg	486	1.27	481	1.44	483	1.12	491	1.50		
Macao (China)	544	1.11	509	1.25	529	1.06	534	1.24		
Malaysia	446	3.25	431	3.48	443	3.00	440	3.29		
Malta	479	1.72	447	1.78	465	1.64				
Mexico	408	2.24	423	2.58	416	2.13	433	2.46		
Moldova	420	2.47	416	2.52	428	1.97				
Montenegro	418	1.46	427	1.58	411	1.03	416	1.27		
Netherlands	512	2.21	503	2.41	509	2.26	518	2.39	509	3.32
New Zealand	495	2.27	509	2.40	513	2.38	533	2.45		
Norway	502	2.23	513	2.51	498	2.26	502	2.52		
Peru	387	2.71	398	2.89	397	2.36	418	2.50	403	3.40
Poland	504	2.39	506	2.48	501	2.51			485	2.97
Portugal	492	2.49	498	2.69	501	2.43	498	2.64		
Qatar	402	1.27	402	1.02	418	1.00				
Romania	444	3.79	434	4.07	435	3.23				
Russia	494	3.11	495	3.08	487	2.91	473	3.42	512	3.33
Singapore	564	1.47	535	1.63	556	1.20	561	1.21		
Slovak Republic	475	2.66	453	2.83	461	2.59	463	2.38	445	4.53
Slovenia	510	1.26	505	1.47	513	1.32	502	1.75		
Spain	486	2.15	496	2.36	493	2.07	496	2.15	469	3.19
Sweden	494	3.17	500	3.48	493	3.60	510	3.44		
Switzerland	521	2.92	492	3.03	506	2.90				
Chinese Taipei	542	3.03	497	2.50	532	2.69	527	2.47		
Thailand	415	3.03	409	3.35	421	2.83	436	3.50		
Trinidad and Tobago	417	1.41	427	1.49	425	1.41				
Tunisia	367	2.95	361	3.06	386	2.10	382	1.94		
Turkey	420	4.13	428	3.96	425	3.93	422	3.45		
United Arab Emirates	427	2.41	434	2.87	437	2.42	435	2.43		
United Kingdom	492	2.50	498	2.77	509	2.56	519	2.68		
United States	470	3.17	497	3.41	496	3.18	520	3.64	487	3.80
Uruguay	418	2.50	437	2.55	435	2.20	443	2.29		
Viet Nam	495	4.46	487	3.73	525	3.91				

See note 2 under Table 12.5.



LINKING ERROR

An evaluation of the magnitude of linking error can be accomplished by considering differences between reported country results from previous PISA cycles and the transformed results from the rescaling. In the application to linking error estimation for the 2015 PISA trend comparisons the robust measure of standard deviation was used, the S_n statistic (Rousseeuw & Croux, 1993); see Chapter 9 for more information on the linking error approach taken in PISA 2015. The robust estimates of linking error between cycles, by domain are presented in Table 12.8.

The S_n statistic is available in SAS as well as the R package robustbase. See also <https://cran.r-project.org/web/packages/robustbase/robustbase.pdf>. The S_n statistic was proposed by Rousseeuw and Croux (1993) as a more efficient alternative to the scaled median absolute deviation from the median ($1.4826 \times \text{MAD}$) that is commonly used as a robust estimator of standard deviation.

Table 12.8 Robust link error (based on absolute pairwise differences statistic S_n) for comparisons of performance between PISA 2015 and previous assessments

Comparison	Maths	Reading	Science	Financial literacy
PISA 2000 to 2015		6.8044		
PISA 2003 to 2015	5.6080	5.3907		
PISA 2006 to 2015	3.5111	6.6064	4.4821	
PISA 2009 to 2015	3.7853	3.4301	4.5016	
PISA 2012 to 2015	3.5462	5.2535	3.9228	5.3309

Note: Comparisons between PISA 2015 scores and previous assessments can only be made to when the subject first became a major domain. As a result, comparisons in mathematics performance between PISA 2015 and PISA 2000 are not possible, nor are comparisons in science performance between PISA 2015 and PISA 2000 or PISA 2003.

INTERNATIONAL CHARACTERISTICS OF THE ITEM POOL

This section provides an overview of the test targeting, the domain inter-correlations and the correlations among the science subscales.

Test targeting

In addition to identifying the relative discrimination and difficulty of items, IRT can be used to summarise the results for various subpopulations of students. A specific value – the response probability (RP) – can be assigned to each item on a scale according to its discrimination and difficulty, similar to students who receive a specific score along a scale according to their performance on the assessment items (OECD, 2002). Chapter 15 describes how items can be placed along a scale based on RP values and how these values can be used to describe different proficiency levels.

After the estimation of item parameters in the item calibration stage, RP values were calculated for each item, and then items were classified into proficiency levels within the cognitive domain. Likewise, after generation of the plausible values, respondents can be classified into proficiency levels for each cognitive domain. The purpose of classifying items and respondents into levels is to provide more descriptive information about group proficiencies. The different item levels provide information about the underlying characteristics of an item as it relates to the domain (such as item difficulty); the higher the difficulty, the higher the level. In PISA, an RP62 value is used for the classification of items into levels. Respondents with a proficiency located below this point have a lower probability than the chosen RP62 value, and respondents with a proficiency above this point have a higher probability (that is > 0.62) of solving an item. The RP62 values for all items are presented in Annex A together with the final item parameters obtained from the IRT scaling. The respondent classification into different levels is done by PISA scale scores transformed from the plausible values. Each level is defined by certain score boundaries for each cognitive domain. Tables 12.9 to 12.13 show the score boundaries overall countries used for each cognitive domain along with the percentage of items and respondents classified at each level of proficiency. The decision for the score boundaries for science is explained in Chapter 15; for reading and mathematics the same levels were used that were defined in previous PISA cycles.

Table 12.9 Item and respondent classification for each score boundary in mathematics

Level	Score points on the PISA scale	Number of items	Percentage of items	Percentage of respondents
6	Higher than 669.30	27	13.30	1.91
5	Higher than 606.99 and less than or equal to 669.30	23	11.33	6.37
4	Higher than 544.68 and less than or equal to 606.99	50	24.63	13.93
3	Higher than 482.38 and less than or equal to 544.68	41	20.20	20.16
2	Higher than 420.07 and less than or equal to 482.38	39	19.21	21.81
1	Higher than 357.77 and less than or equal to 420.07	12	5.91	18.78
Below 1	Less than 357.77	11	5.42	17.05

Table 12.10 Item and respondent classification for each score boundary in reading

Level	Score points on the PISA scale	Number of items	Percentage of items	Percentage of respondents
6	Higher than 698.32	18	7.63	0.70
5	Higher than 625.61 and less than or equal to 698.32	28	11.86	4.96
4	Higher than 552.89 and less than or equal to 625.61	50	21.19	15.45
3	Higher than 480.18 and less than or equal to 552.89	62	26.27	24.14
2	Higher than 407.47 and less than or equal to 480.18	58	24.58	24.36
1a	Higher than 334.75 and less than or equal to 407.47	15	6.36	17.92
1b	262.04 to less than or equal to 334.75	5	2.12	9.12
Below 1b	Less than 262.04	0	0.00	3.34

Table 12.11 Item and respondent classification for each score boundary in science

Level	Score points on the PISA scale	Number of items	Percentage of items	Percentage of respondents
6	Higher than 707.93	13	4.45	0.76
5	Higher than 633.33 and less than or equal to 707.93	29	9.93	4.79
4	Higher than 558.73 and less than or equal to 633.33	75	25.68	14.51
3	Higher than 484.14 and less than or equal to 558.73	94	32.19	23.20
2	Higher than 409.54 and less than or equal to 484.14	63	21.58	25.71
1a	Higher than 334.94 and less than or equal to 409.54	15	5.14	20.88
1b	260.54 to less than or equal to 334.94	3	1.03	8.68
Below 1b	Less than 260.54	0	0.00	1.48

Table 12.12 Item and respondent classification for each score boundary in financial literacy

Level	Score points on the PISA scale	Number of items	Percentage of items	Percentage of respondents
5	Higher than 624.63	22	26.51	9.36
4	Higher than 549.86 and less than or equal to 624.63	14	16.87	17.38
3	Higher than 475.10 and less than or equal to 549.86	24	28.92	24.31
2	Higher than 400.33 and less than or equal to 475.10	12	14.46	22.63
1	Higher than 325.57 and less than or equal to 400.33	6	7.23	15.73
Below 1	Less than 325.57	5	6.02	10.59

Table 12.13 Item and respondent classification for each score boundary in CPS

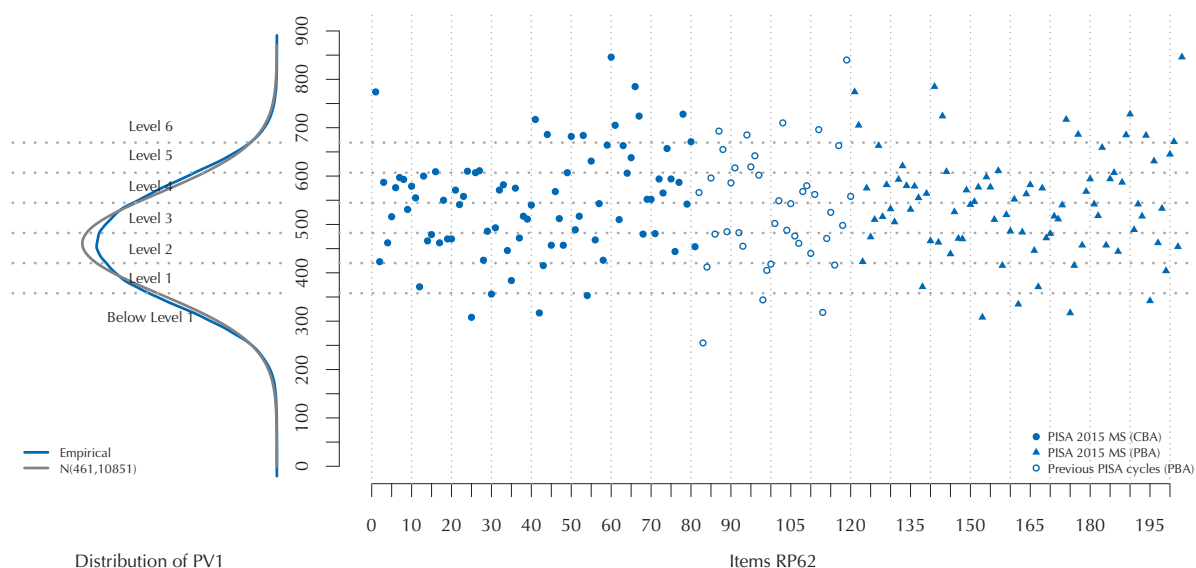
Level	Score points on the PISA scale	Number of items	Percentage of items	Percentage of respondents
4	Higher than 640.00	25	21.37	6.28
3	Higher than 540.00 and less than or equal to 640.00	28	23.93	23.66
2	Higher than 440.00 and less than or equal to 540.00	38	32.48	35.30
1	Higher than 340.00 and less than or equal to 440.00	20	17.09	26.78
Below 1	Less than 340.00	6	5.13	7.99



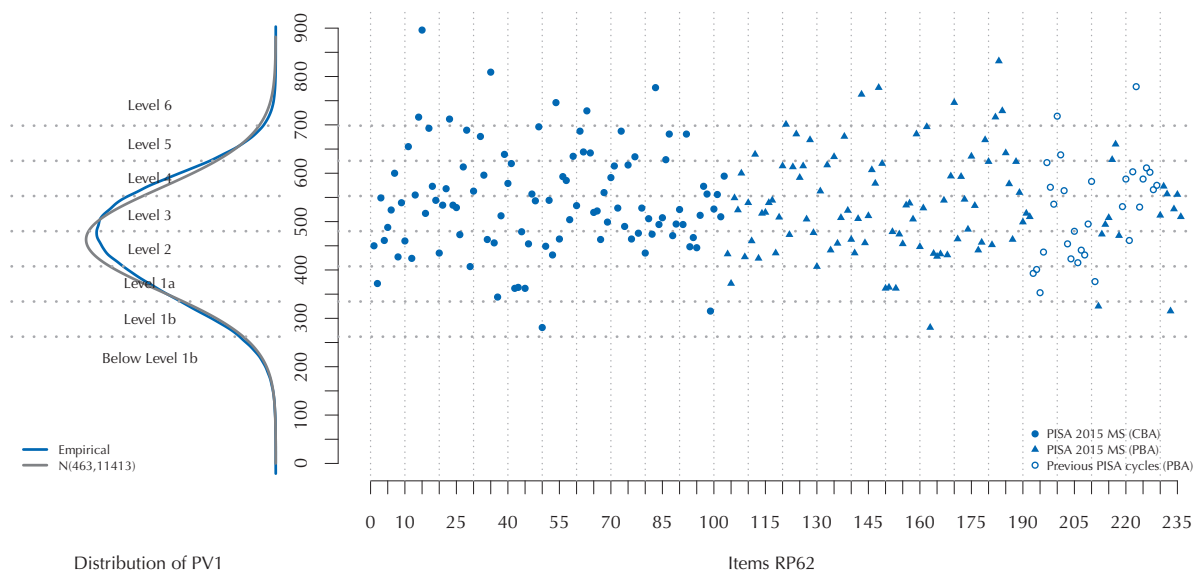
Because RP62 values and the transformed plausible values are on the same PISA scales, the distribution of respondents' latent ability and item RP62 values can be located on the same scale. Figures 12.7 to 12.11 illustrate the distribution of the first plausible value (PV1) along with item RP62 values on the PISA scale separately for each cognitive domain for the PISA 2015 main survey data. Note that international RP62 values and international plausible values (PV1) were used for these figures.¹ RP62 values for CBA items are denoted on the right side. In each domain, solid circles indicate PBA items and hollow circles indicate additional PBA items from previous PISA cycles that were not administered in PISA 2015 main survey. For the polytomous items where partial scoring was available, only the highest RP62 values are illustrated in these figures. On the left side, the distribution of plausible values is plotted. In each figure, the blue line indicates the empirical density of the plausible values across countries, and the grey line indicates the theoretical normal distribution with mean of plausible values and the variance of plausible values in each domain across countries. Specifically, $N(461, 104.17^2)$ for mathematics, $N(463, 106.83^2)$ for reading, $N(467, 103.02^2)$ for science, $N(474, 123^2)$ for financial literacy, and $N(483, 101.65^2)$ for CPS are displayed as grey lines. (Note that there are RP62 values higher than 1 000 for the CPS domain, these are outside of the region occupied by the vast majority of respondent's proficiency estimates and therefore are not shown in Figure 12.11.)

■ Figure 12.7 ■

Item RP62 values and distribution of PV1 in maths



■ Figure 12.8 ■

Item RP62 values and distribution of PV1 in reading

■ Figure 12.9 ■

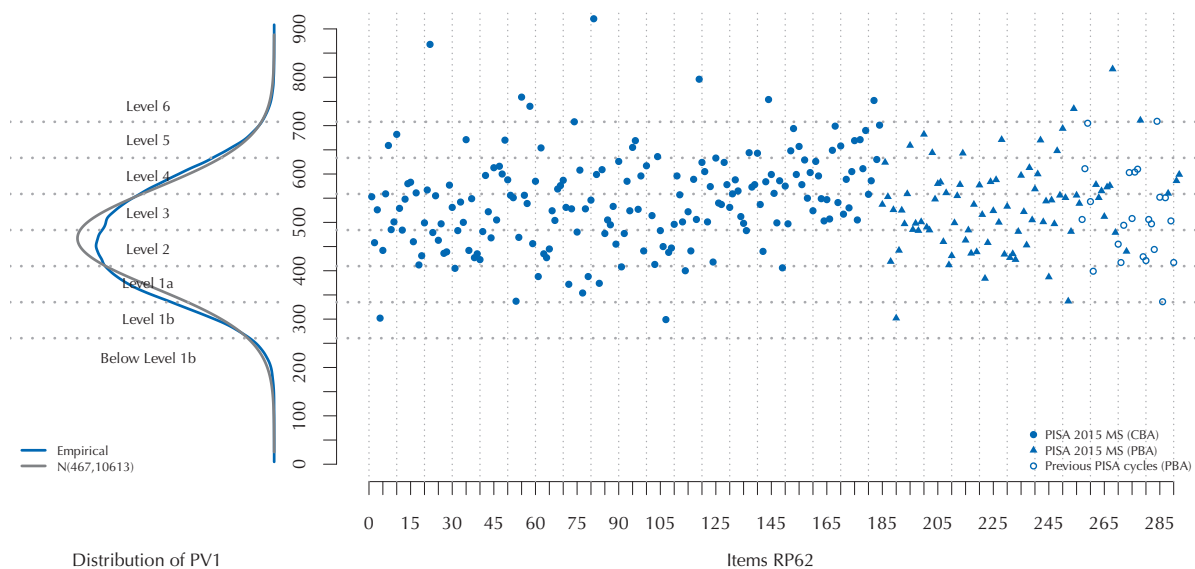
Item RP62 values and distribution of PV1 in science



Figure 12.10

Item RP62 values and distribution of PV1 in financial literacy

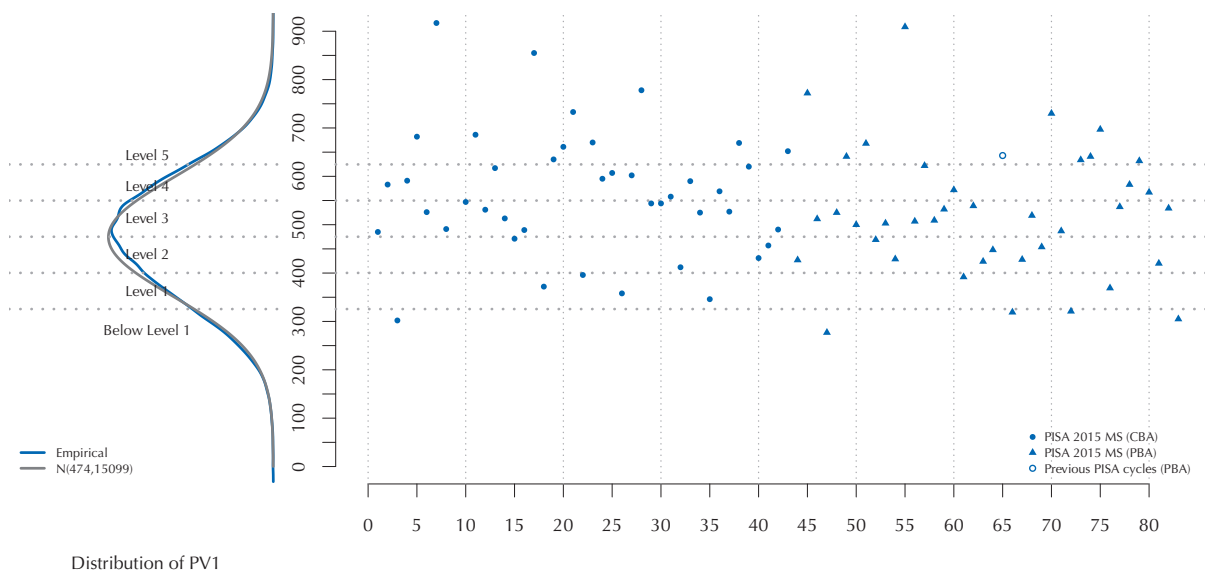
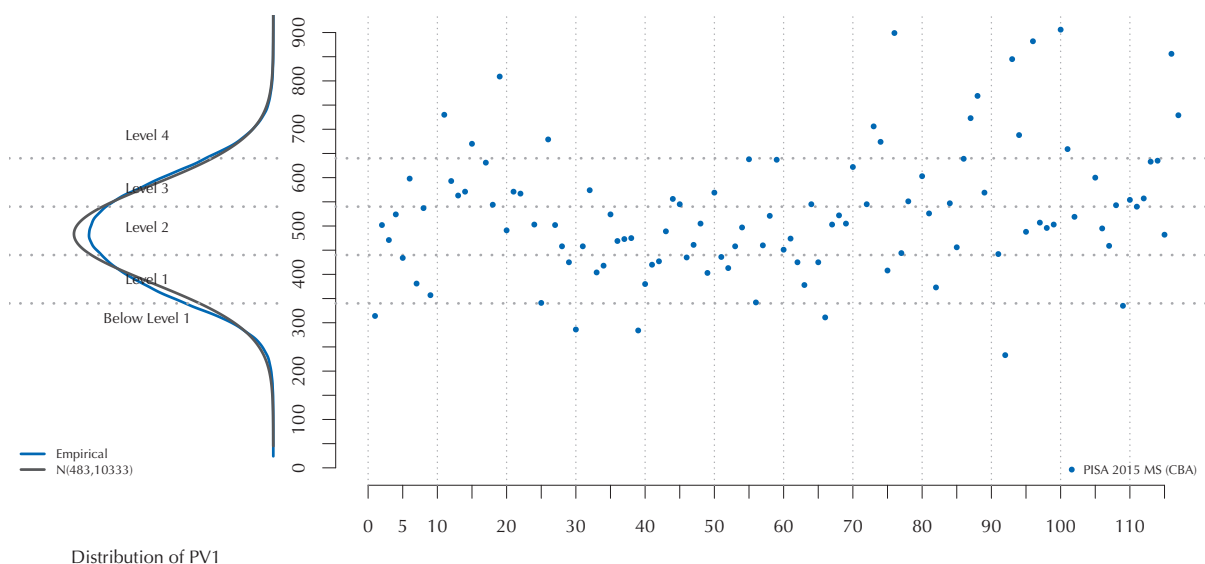


Figure 12.11

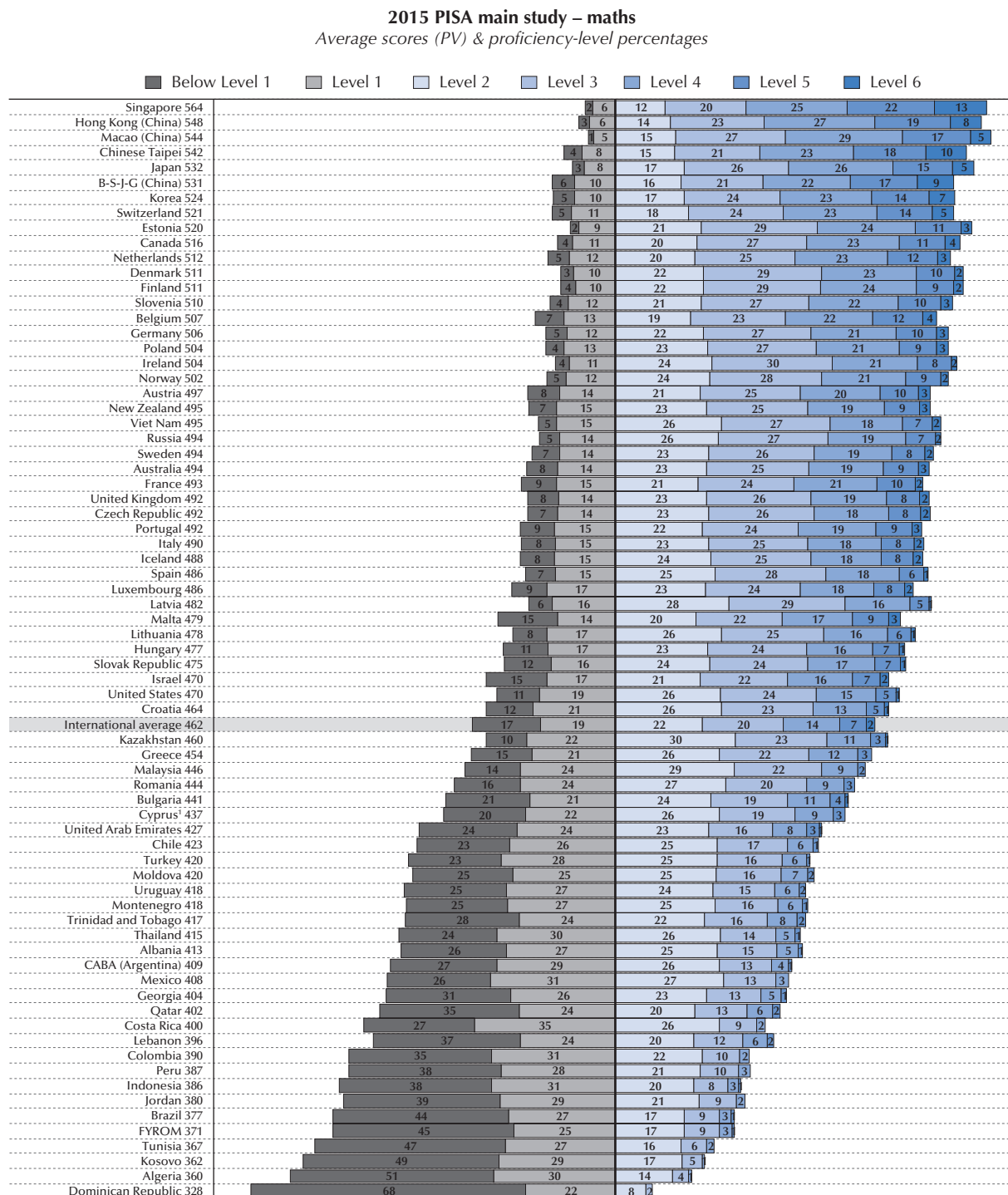
Item RP62 values and distribution of PV1 in collaborative problem solving



Figures 12.12 to 12.16 show the percentage of respondents per country at each level of proficiency for each cognitive domain.

■ Figure 12.12 ■

Percentage of respondents per country/economy at each level of proficiency for maths



1. 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 recognizes 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."

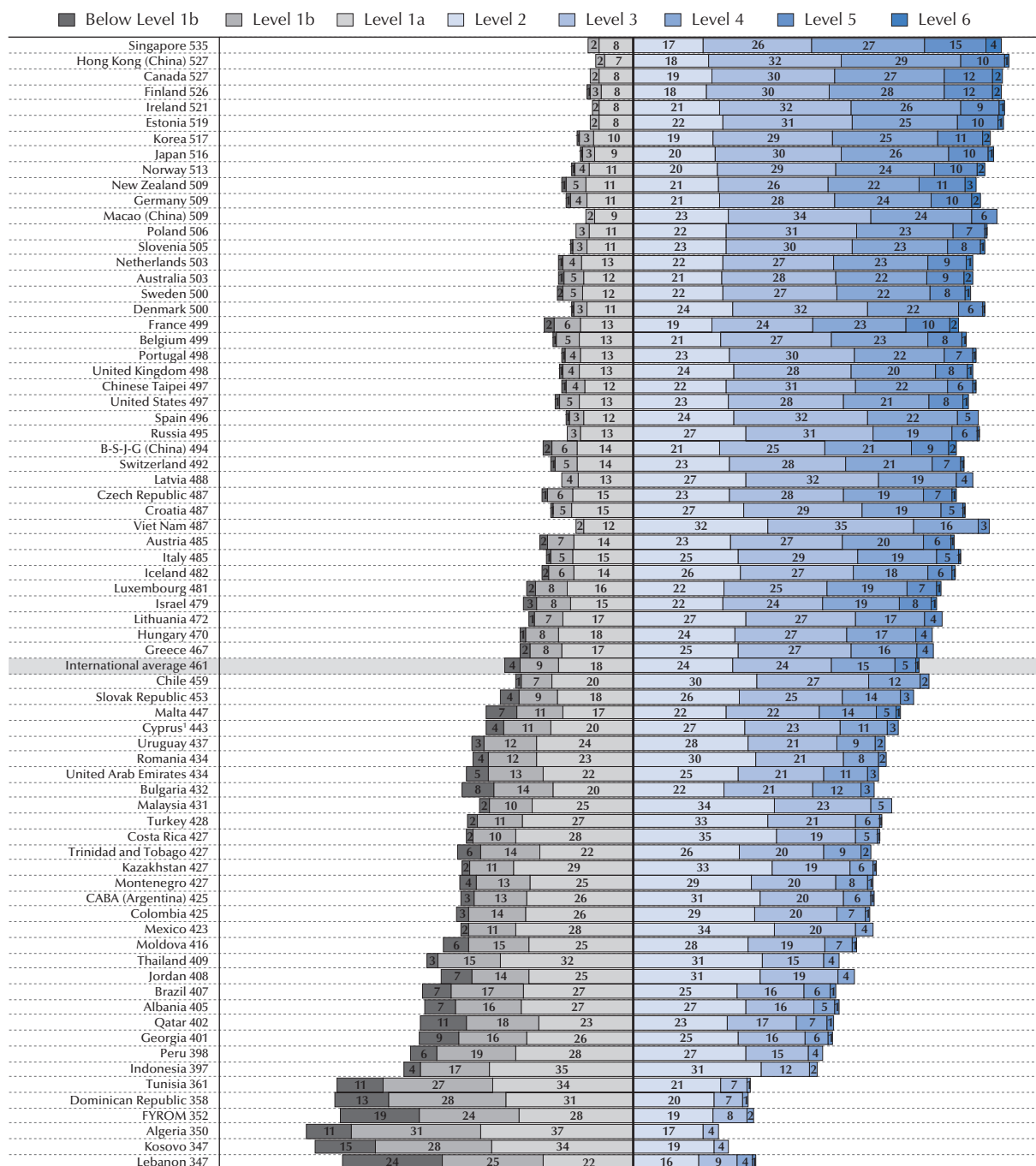
Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of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



■ Figure 12.13 ■

Percentage of respondents per country/economy at each level of proficiency for reading

2015 PISA main study – reading
Average scores (PV) & proficiency-level percentages



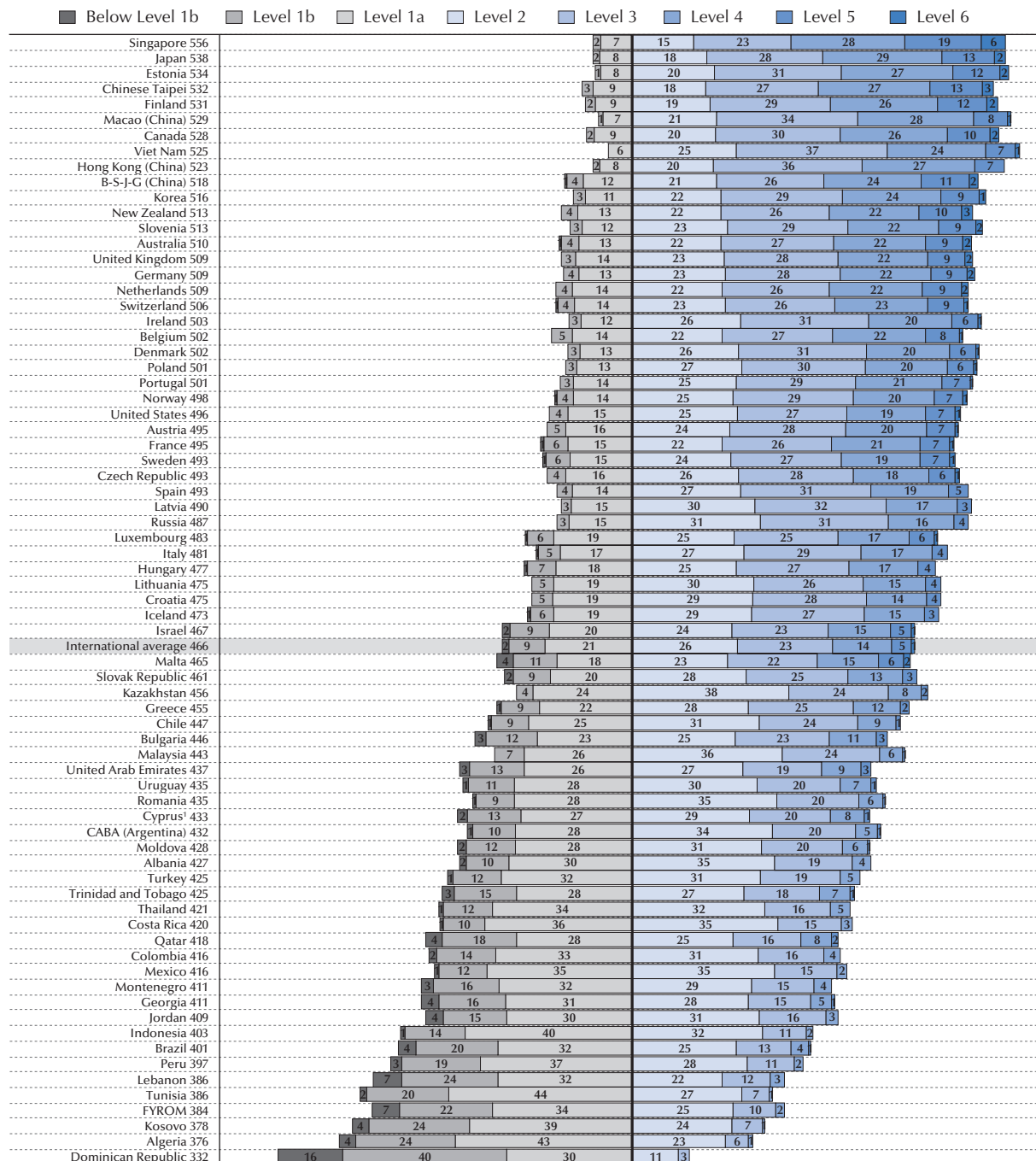
1. See note 2 under Table 12.5.

■ Figure 12.14 ■

Percentage of respondents per country/economy at each level of proficiency for science

2015 PISA main study – science

Average scores (PV) & proficiency-level percentages



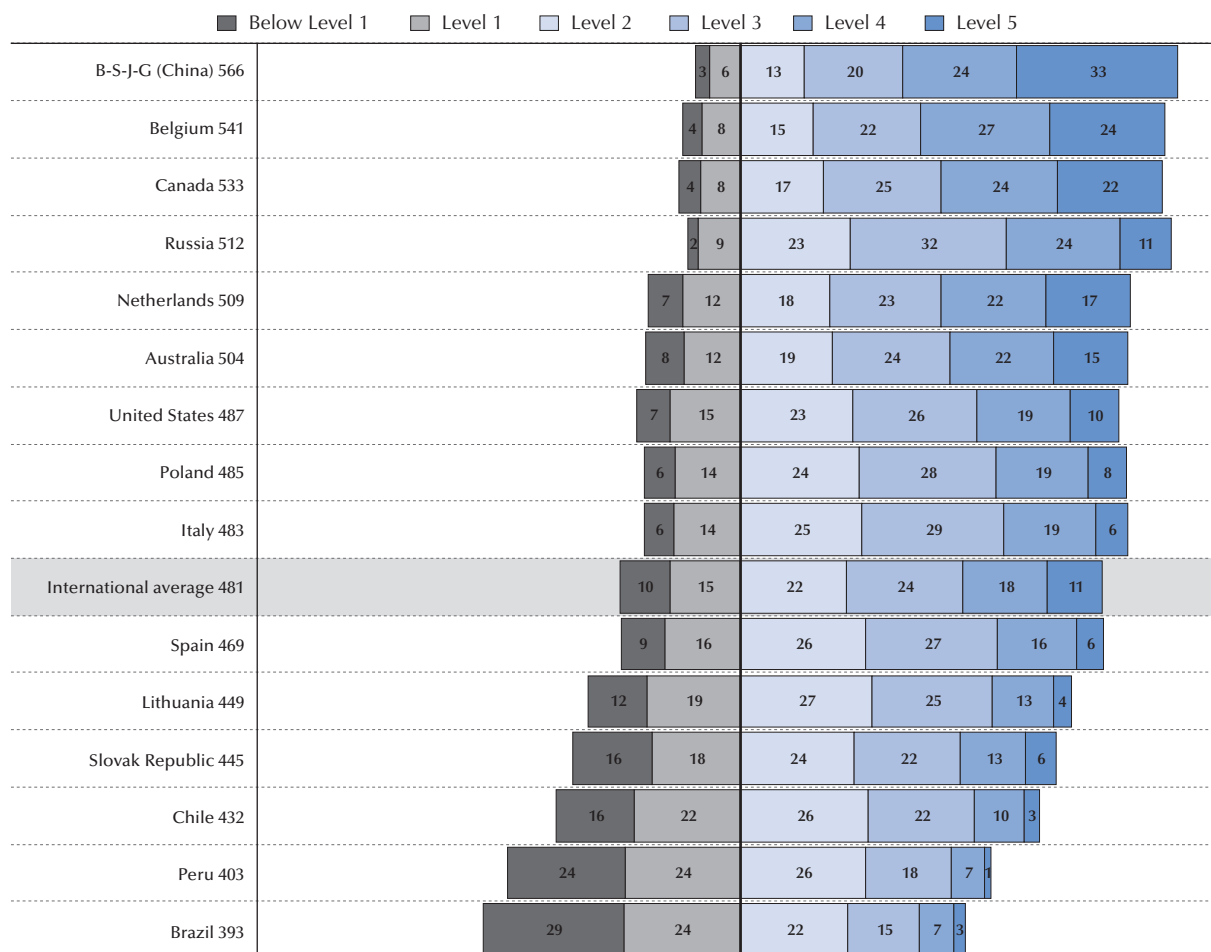
1. See note 2 under Table 12.5.



■ Figure 12.15 ■

Percentage of respondents per country/economy at each level of proficiency for financial literacy

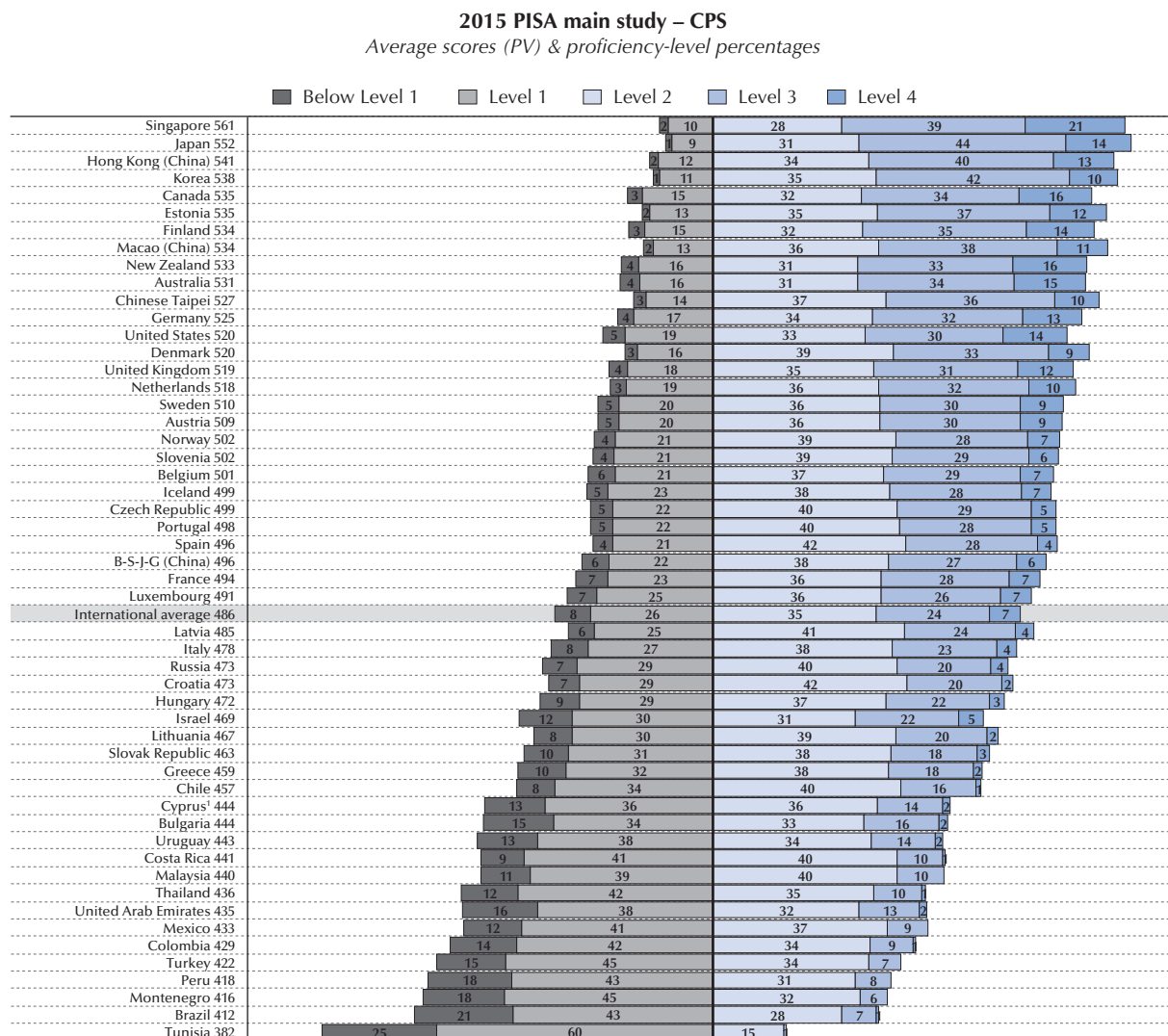
2015 PISA main study – financial literacy
Literacy average scores (PV) & proficiency-level percentages



Note: The financial literacy data from Belgium come from the Flanders part of Belgium only and thus are not nationally representative; the same is the case with regard to the financial literacy data from Canada since some provinces of Canada did not participate in the financial literacy assessment.

■ Figure 12.16 ■

Percentage of respondents per country/economy at each level of proficiency for CPS



Note: The CPS sample from Israel does not include ultra-Orthodox students and thus is not nationally representative.

1. See note 2 under Table 12.5.

Domain inter-correlations

Estimated correlations between the PISA domains, based on the 10 plausible values and averaged across all countries and assessment modes, are presented in Table 12.14. Overall, the correlations are quite high, as expected, yet there is still some separation between each of the domains. The estimated correlations at the national level are presented in Table 12.15.

Table 12.14 Domain inter-correlations¹

Domain		Reading	Science	CPS	Financial literacy
Maths	Average	0.79	0.88	0.70	0.74
	Average (CBA)	0.79	0.88	0.70	0.74
	Average (PBA)	0.75	0.80	–	–
	Range	0.57–0.87	0.70–0.91	0.55–0.76	0.60–0.81
Reading	Average	–	0.87	0.74	0.75
	Average (CBA)		0.87	0.74	0.75
	Average (PBA)		0.77	–	–
	Range		0.71–0.90	0.58–0.80	0.61–0.81
Science	Average	–	–	0.77	0.77
	Average (CBA)			0.77	0.77
	Average (PBA)			–	–
	Range			0.65–0.83	0.68–0.85
CPS	Average	–	–	–	0.64
	Average (CBA)				0.64
	Average (PBA)				–
	Range				0.50–0.71

1. Please note that Argentina, Malaysia and Kazakhstan were not included in this analysis due to adjudication issues (inadequate coverage of either population or construct).

[Part 1/2]

Table 12.15 National-level domain inter-correlations based on 10 PVs

Country/economy	Maths & reading	Maths & science	Maths & CPS	Maths & fin. lit.	Reading & science	Reading & CPS	Reading & fin. lit.	Science & CPS	Science & fin. lit.	CPS & fin. lit.
Albania	0.68	0.80	–	–	0.77	–	–	–	–	–
Algeria	0.57	0.70	–	–	0.71	–	–	–	–	–
Argentina	0.75	0.83	–	–	0.81	–	–	–	–	–
Australia	0.79	0.88	0.68	0.79	0.87	0.75	0.8	0.76	0.85	0.7
Austria	0.80	0.89	0.71	–	0.88	0.77	–	0.78	–	–
B-S-J-G (China)	0.84	0.91	0.76	0.80	0.90	0.76	0.80	0.80	0.83	0.70
Belgium	0.84	0.90	0.73	0.80	0.90	0.76	0.80	0.78	0.83	0.67
Brazil	0.75	0.84	0.65	0.62	0.86	0.73	0.65	0.75	0.68	0.54
Bulgaria	0.80	0.89	0.74	–	0.89	0.80	–	0.83	–	–
Canada	0.77	0.87	0.67	0.68	0.87	0.74	0.70	0.75	0.74	0.59
Chile	0.80	0.88	0.70	0.75	0.87	0.74	0.75	0.77	0.78	0.64
Colombia	0.83	0.90	0.74	–	0.90	0.74	–	0.80	–	–
Costa Rica	0.75	0.83	0.59	–	0.85	0.67	–	0.68	–	–
Croatia	0.80	0.89	0.69	–	0.87	0.75	–	0.76	–	–
Cyprus ¹	0.74	0.85	0.65	–	0.83	0.71	–	0.74	–	–
Czech Republic	0.84	0.90	0.69	–	0.89	0.72	–	0.75	–	–
Denmark	0.77	0.87	0.69	–	0.86	0.72	–	0.77	–	–
Dominican Republic	0.78	0.83	–	–	0.85	–	–	–	–	–
Estonia	0.78	0.88	0.71	–	0.87	0.74	–	0.79	–	–
Finland	0.79	0.87	0.72	–	0.87	0.75	–	0.78	–	–
France	0.84	0.91	0.70	–	0.90	0.75	–	0.78	–	–
FYROM	0.75	0.78	–	–	0.74	–	–	–	–	–
Georgia	0.79	0.79	–	–	0.73	–	–	–	–	–
Germany	0.81	0.90	0.70	–	0.88	0.72	–	0.77	–	–
Greece	0.79	0.88	0.73	–	0.88	0.75	–	0.79	–	–
Hong Kong	0.77	0.88	0.64	–	0.86	0.73	–	0.74	–	–
Hungary	0.83	0.90	0.74	–	0.90	0.78	–	0.81	–	–
Iceland	0.78	0.86	0.70	–	0.84	0.74	–	0.76	–	–
Indonesia	0.70	0.82	–	–	0.75	–	–	–	–	–
Ireland	0.81	0.89	–	–	0.88	–	–	–	–	–
Israel	0.83	0.89	0.75	–	0.89	0.78	–	0.80	–	–

[Part 2/2]

Table 12.15 National-level domain inter-correlations based on 10 PVs

Countries	Maths & Reading	Maths & Science	Maths & CPS	Maths & Fin. Lit.	Reading & Science	Reading & CPS	Reading & Fin. Lit.	Science & CPS	Science & Fin. Lit.	CPS & Fin. Lit.
Italy	0.75	0.85	0.65	0.68	0.84	0.68	0.67	0.73	0.73	0.56
Japan	0.79	0.87	0.66	–	0.86	0.73	–	0.72	–	–
Jordan	0.70	0.79	–	–	0.78	–	–	–	–	–
Kazakhstan	0.61	0.73	–	–	0.70	–	–	–	–	–
Korea	0.78	0.87	0.72	–	0.85	0.76	–	0.77	–	–
Kosovo	0.74	0.81	–	–	0.78	–	–	–	–	–
Latvia	0.77	0.87	0.66	–	0.87	0.73	–	0.75	–	–
Lebanon	0.80	0.82	–	–	0.81	–	–	–	–	–
Lithuania	0.79	0.90	0.72	0.70	0.87	0.74	0.73	0.79	0.75	0.63
Luxembourg	0.83	0.91	0.73	–	0.90	0.78	–	0.78	–	–
Macao	0.75	0.84	0.65	–	0.89	0.78	–	0.78	–	–
Malaysia	0.78	0.87	0.72	–	0.88	0.74	–	0.79	–	–
Malta	0.83	0.87	–	–	0.87	–	–	–	–	–
Mexico	0.77	0.84	0.67	–	0.86	0.73	–	0.76	–	–
Moldova	0.73	0.79	–	–	0.77	–	–	–	–	–
Montenegro	0.76	0.83	0.66	–	0.84	0.70	–	0.74	–	–
Netherlands	0.87	0.91	0.75	0.81	0.89	0.78	0.81	0.77	0.84	0.70
New Zealand	0.79	0.89	0.70	–	0.87	0.75	–	0.78	–	–
Norway	0.78	0.89	0.68	–	0.84	0.72	–	0.74	–	–
Peru	0.81	0.86	0.73	0.76	0.88	0.78	0.81	0.79	0.79	0.70
Poland	0.80	0.90	–	0.74	0.86	–	0.75	–	0.77	–
Portugal	0.79	0.89	0.70	–	0.86	0.74	–	0.76	–	–
Qatar	0.84	0.88	–	–	0.90	–	–	–	–	–
Romania	0.79	0.78	–	–	0.77	–	–	–	–	–
Russian Federation	0.66	0.82	0.55	0.60	0.81	0.68	0.61	0.70	0.68	0.50
Singapore	0.82	0.89	0.73	–	0.90	0.78	–	0.80	–	–
Slovak Republic	0.83	0.88	0.69	0.66	0.87	0.74	0.66	0.74	0.68	0.58
Slovenia	0.79	0.89	0.68	–	0.87	0.73	–	0.74	–	–
Spain	0.76	0.88	0.66	0.71	0.86	0.71	0.72	0.74	0.75	0.61
Sweden	0.78	0.89	0.71	–	0.85	0.78	–	0.77	–	–
Switzerland	0.81	0.88	–	–	0.88	–	–	–	–	–
Chinese Taipei	0.83	0.90	0.71	–	0.90	0.77	–	0.77	–	–
Thailand	0.75	0.83	0.65	–	0.87	0.76	–	0.78	–	–
Trinidad and Tobago	0.81	0.87	–	–	0.80	–	–	–	–	–
Tunisia	0.72	0.81	0.59	–	0.83	0.58	–	0.65	–	–
Turkey	0.76	0.86	0.68	–	0.85	0.71	–	0.76	–	–
United Arab Emirates	0.81	0.88	0.74	–	0.89	0.80	–	0.81	–	–
United Kingdom	0.77	0.87	0.68	–	0.86	0.74	–	0.76	–	–
United States	0.83	0.90	0.76	0.80	0.90	0.79	0.80	0.82	0.83	0.71
Uruguay	0.79	0.88	0.71	–	0.87	0.73	–	0.77	–	–
Viet Nam	0.81	0.87	–	–	0.85	–	–	–	–	–

1. See note 2 under Table 12.5.

Science scale and subscales

The estimated correlations between the PISA 2015 science subscales and the domains of reading, mathematics, science and financial literacy scales, are presented in Tables 12.16 to 12.18. The different science subscales, which belong to the three scales or subscale groups Knowledge (SKCO, SKPE), Competency (SCEP, SCED, SCID), and System (SSPH, SSLI, SSES), were considered.

Please note that because of the way in which the proficiency data were generated, you should not calculate the correlations among the knowledge, competency and systems subscales. Therefore these are presented in separate tables.

**Table 12.16 Estimated correlations among domains and science knowledge subscales¹**

	Reading	Science	CPS	Financial literacy	SKCO	SKPE
Maths	0.783	0.863	0.692	0.726	0.798	0.808
Reading		0.853	0.741	0.738	0.786	0.817
Science			0.765	0.770	–	–
CPS				0.630	0.688	0.722
FinLit					0.743	0.763
SKCO						0.921

Note: Content, SKPE: Procedural & Epistemic.

1. Please note that Argentina, Malaysia and Kazakhstan were not included in this analysis due to adjudication issues (inadequate coverage of either population or construct).

Table 12.17 Estimated correlations among domains and science Competency subscales¹

	Reading	Science	CPS	Financial literacy	SCED	SCEP	SCID
Maths	0.783	0.863	0.692	0.726	0.778	0.797	0.802
Reading		0.853	0.741	0.738	0.790	0.786	0.805
Science			0.765	0.770	–	–	–
CPS				0.630	0.700	0.687	0.712
FinLit					0.733	0.743	0.756
SCED						0.894	0.903
SCEP							0.919

Note: SCED: Evaluate and Design Scientific Inquiry, SCEP: Subscale of Science Explain Phenomena Scientifically, SCID: Interpret Data and Evidence Scientifically.

1. Please note that Argentina, Malaysia and Kazakhstan were not included in this analysis due to adjudication issues (inadequate coverage of either population or construct).

Table 12.18 Estimated correlations among domains and science System subscales¹

	Reading	Science	CPS	Financial literacy	SSES	SSLI	SSPH
Maths	0.783	0.863	0.692	0.726	0.791	0.798	0.791
Reading		0.853	0.741	0.738	0.791	0.804	0.781
Science			0.765	0.770	---	---	---
CPS				0.630	0.693	0.711	0.688
FinLit					0.743	0.754	0.736
SSES						0.910	0.900
SSLI							0.908

Note: SSPH: Physical, SSLI: Living, SSES: Earth & Science.

1. Please note that Argentina, Malaysia and Kazakhstan were not included in this analysis due to adjudication issues (inadequate coverage of either population or construct).



Note

1. Please note that Argentina, Malaysia and Kazakhstan were not included in this analysis due to adjudication issues (inadequate coverage of either population or construct).

References

Efron, B. (1982), "The Jackknife, the Bootstrap, and Other Resampling Plans", *Society of Industrial and Applied Mathematics CBMS-NSF Monographs*, Vol. 38.

Hoaglin, D.C., F. Mosteller and J.W. Tukey, (1983), *Understanding Robust and Exploratory Data Analysis*, John Wiley & Sons, New York, NY.

Mislevy, R.J. and K.M. Sheehan, (1987), "Marginal estimation procedures", in A.E. Beaton (Ed.), *Implementing the new design: The NAEP 1983-84 technical report*, (Report No. 15-TR-20), Educational Testing Service, Princeton, NJ.

OECD (2002), *Reading for Change: Performance and Engagement across Countries: Results from PISA 2000*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264099289-en>.

von Davier et al. (2006), "The statistical procedures used in National Assessment of Educational Progress: Recent developments and future directions", in C.R. Rao and S. Sinharay (Ed.), *Handbook of Statistics*, Vol. 26, pp. 1039-1055, Elsevier.