17

Validation of the Embedded Attitudinal Scales

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

INTRODUCTION

The development processes that are employed by PISA to ensure the cross-national validity of its scales consist of four steps. First, the construct should have well-established theoretical underpinnings. That is the construct should be underpinned by a body of academic literature and it should be supported by leading theorists and academics working in an area. Within PISA this is ensured through the articulation of the constructs in widely discussed and reviewed assessment frameworks. For the *embedded interest* and *embedded support* scales the articulation can be found in the *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006* (OECD 2006) (also see Chapter 2).

Second, the approach to operationalising the construct must be widely supported – that is, there must be wide agreement that the items that are used in PISA are reflective of the underlying conceptual definition of the domain. For the *embedded interest* and *embedded support* scales the procedures used in PISA to ensure this parallel those used for the cognitive assessment items. The procedures that PISA puts in place to achieve this include:

- The use of skilled professional test development teams from a variety of PISA participating countries;
- Review of the items as they are prepared by experts who have been directly involved in and were often responsible for the conceptualisation of the underpinning construct definitions;
- Opportunities for review and evaluation of the drafted items by PISA participating countries on multiple occasions;
- A detailed set of translation and translation verification protocols that are aimed at ensuring the conceptual and psychometric equivalence of the items across languages and cultures;
- A range of small, medium and large trial testing activities where students are asked to respond to the item and to reflect upon the meaning of the items to them.

Third, psychometric analyses are undertaken to ensure that the sets of items that are deemed to be reflective of the underlying construct can indeed be brought together in a coherent fashion to provide indicators of the underlying construct. These analyses pay particular attention to the scalability, reliability and cross-country consistency of the behaviour of the items.

Finally, the constructed scales are reviewed for their nomothetic span. That is, the extent to which relations with other variables make conceptual sense.

This chapter is concerned with the range of analyses that were undertaken as part of the last two steps in the above-described process for validating the scales that were constructed from the attitudinal items. The purpose of these analyses was to confirm the empirical validity of the scales for the purposes of cross-national comparisons.

For the main study, attitudinal items were embedded within units of the science test in order to obtain measures of two attitudinal dimensions (or constructs): *interest in science* and *support for scientific inquiry*.[1] In short, these domains will be referred to as *embedded interest* and *embedded support.*

As the analyses reported here were undertaken for validation purposes, prior to the finalisation of the international database, they were undertaken with data that had not been fully cleaned and weighted. The majority of the analyses reported use data from 51 different data sets – this was made up of all 30 OECD countries and 21 partner countries. Where this is not the case it is noted.

The thirty OECD datasets were used for the analyses of scale reliability, gender DIF, general confirmation of the expected dimensional structure for embedded science attitude items and the correlation between scales.[2] Random sub-samples of 5000 cases were taken for countries that used over-sampling in the main study. In particular, reduced samples were used for Australia, Belgium, Canada, Italy, Mexico, Spain and Switzerland. The UH booklet responses were excluded from these analyses.[3]

For the item response theory analyses, a calibration sample of 500 cases from each of the OECD datasets was used to estimate item parameters. These item parameter estimates were then used to estimate weighted likelihood estimates for each case for each of the five test scales (*mathematics, reading, science, embedded interest* and *embedded support*).

Preliminary weights were available for all countries except Australia and USA at the time of analysing the data for this report.

INTERNATIONAL SCALABILITY

Analysis of item dimensionality with exploratory and confirmatory factor analysis

Software packages *Mplus* (Muthén and Muthén, 2004) and ACER *ConQuest*® (Wu, Adams and Wilson, 1997) were used to confirm the two dimensional structure of the embedded attitudinal measures (*embedded interest* and *embedded support*).

When the items have Likert-type response categories it is recommended that factor analyses should be conducted on the matrix of polychoric inter-item correlations rather than on the matrix of product-moment correlations. Unfortunately, exploratory factor analyses (EFA) based on polychoric correlations has only been implemented in the software package *Mplus* for complete datasets. Because PISA uses a rotated booklet design, EFA was undertaken with *Mplus* with the variables defined as continuous and with product-moment correlations.[4]

Appendix 7 gives the *Mplus* results for both an EFA (with a *promax* rotation) and a two dimensional confirmatory factor analysis (CFA). The results can be summarised as follows:

- *Interest in learning science (embedded interest)*: Solutions with two factors confirmed that interest items generally loaded on one dimension. However, some items (S456N-THE CHEETAH, S519N-AIRBAGS and S527N-EXTINCTION OF THE DINOSAURS) were loading on the second factor – that is the *support* factor;
- *Support for scientific inquiry (embedded support)*: the items selected for main study for this domain items loaded on one factor;
- For the CFA the estimated latent correlation between *embedded interest* and *embedded support* was 0.594;
- The RMSEA measure of model fit produced by *Mplus* was 0.025, which was considered quite acceptable.

Fit to item response model

An alternative approach to assessing item dimensionality is to assess the fit of the data to a multi-dimensional IRT model. Here, a five-dimensional model (reading, mathematics, science, embedded interest and embedded support) was fit to the data using the *ConQuest*® software (Wu, Adams and Wilson, 1997).

The item-level fit statistics for each of the attitudinal items is given in Appendix 7, a normal probability plot of the fit mean squares is given in Figure 17.1 and the estimated latent correlations for the five-dimensional IRT model are given in Table 17.1. The normal probability plot provides a comparison of the distribution of the fit statistics with normal distribution that would be expected if the data did fit the model.

The range and distributions of the fit statistics show an acceptable fit to the multi-dimensional item response model. The fit mean squares are close to normally distributed and the worst fit mean square is 1.18.

Figure 17.1
Distribution of item fit mean square statistics for embedded attitude items

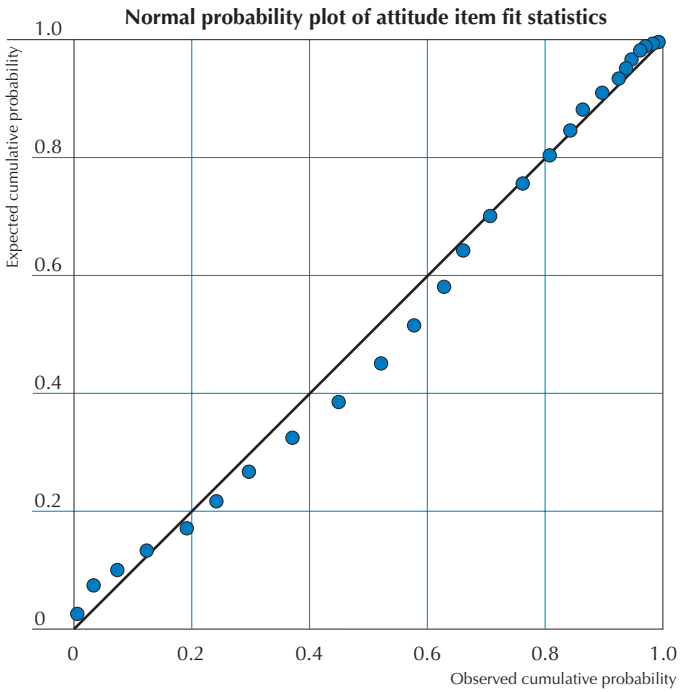


Table 17.1 shows that the estimated latent correlation between *embedded interest* and *embedded support* is 0.623. This is very similar to the corresponding value, 0.594, that was estimated using the CFA.

The correlation of the *embedded support* scale with the three achievement scales is about 0.20, while the correlation between the *embedded interest* scale with the three achievement scales is about –0.15.[5] We return to explore this negative correlation, at the student level, later in this chapter.

Table 17.1
Student-level latent correlations between mathematics, reading, science, embedded interest and embedded support

	Mathematics	Reading	Science	Embedded Interest
Reading	0.780			
Science	0.871	0.831		
Embedded Interest	–0.194	–0.151	–0.133	
Embedded Support	0.136	0.215	0.223	0.623

Reliability

Further scaling properties of the embedded items are reported in Chapter 13, where the overall reliability of the *embedded interest* scale is estimated as 0.892, and of the *embedded support* scale 0.818 (using WLEs). The reliabilities by country are also reported in Chapter 12.

At the country level the reliabilities are greater than 0.80 for *embedded interest* and greater than 0.70 for the *embedded support* scale. As discussed in Chapter 12 the lower reliability for the *embedded support* scale is likely due to the fact that the majority of student responded positively to the support items – *i.e.* students overwhelming expressed positive support for science.

Differential item functioning

Country DIF

IRT models were also estimated for each country data set separately. Comparing the outcomes with the results for the pooled international sample (51 countries) provides information about potential item-by-country interactions, which is a case of differential item function (DIF) associated with the country of test. In addition, it is informative to review item discrimination and item fit statistics in order to assess whether the scaling model holds across countries.

Table 17.2
Summary of the IRT scaling results across countries

	Item-by-country Interaction			Number of items with a discrimination < 0.20	Weighted MNSQ Fit	
	Number of items easier than expected	Number of items harder than expected	Number of items with country DIF		Number of items with fit < 0.8	Number of items with fit > 1.2
For interest in...						
No countries	26	32	13	52	52	42
1 or 2 countries	22	8	22	0	0	4
3 countries or more	4	12	17	0	0	6
N (items)	52	52	52	52	52	52
For support in...						
No countries	25	22	12	37	37	33
1 or 2 countries	11	12	20	0	0	3
3 countries or more	1	3	5	0	0	1
N (items)	37	37	37	37	37	37

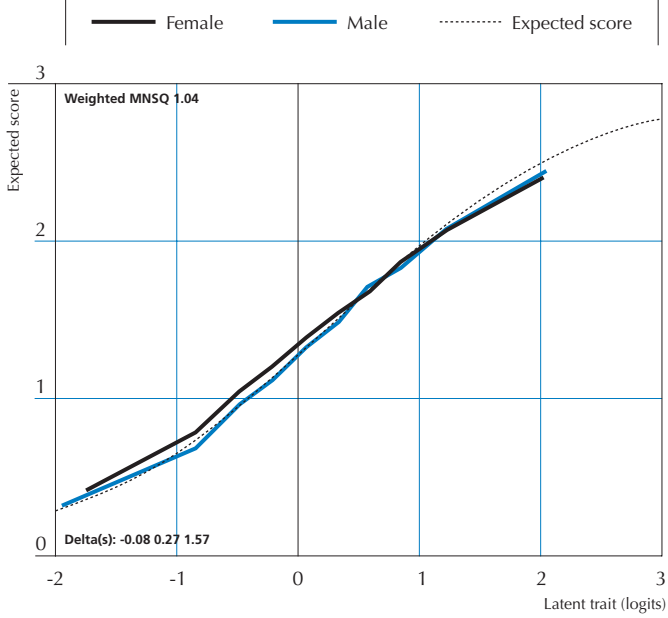
Table 17.2 summarises the results of the national scaling analyses. For each attitude scale it shows the number of items that were significantly easier, harder, or different (easier or harder) compared to the pooled international sample in the following categories: (i) in no country, (ii) in only one or two countries or (iii) in three countries or more. The fourth column gives the number of items with low discrimination (item-score correlations below 0.20), the fifth column the number of items with a weighted MNSQ item fit lower than 0.8 or higher than 1.2 in each of the categories described above.

Item-by-country interactions indicate the degree of parameter invariance across countries. The results show that parameters for items measuring both *embedded interest* and *embedded support* tend to be fairly stable across countries. Furthermore, there were no items with a discrimination value less than 0.20. A full set of item-by-country interaction plots for the embedded items and cognitive items was constructed. An analysis of the item-by-country interactions shows that the embedded item parameter estimates are more stable across countries than the parameter estimates for the cognitive items.

Gender DIF

To investigate any effect of gender DIF on item performance, Expected Score Curves (ESC) were constructed and reported. A full set of plots showing the gender DIF for each embedded item was constructed. Figure 17.2 shows an example of ESC plots for item S408RNA (S408QNA recoded so that strongly agree = 3, agree = 2, disagree = 1, strongly disagree = 0). The solid line represents a predicted score and dots are observed scores for females and males separately.

Figure 17.2
An example of the ESC plot for item S408RNA



The gender DIF analysis was performed by one run of a multi-facet Rasch model, where item difficulty was modelled as a function of item, gender and item-by-gender interaction terms. Table 17.3 shows a tabular report of the gender DIF for embedded attitude items. For each item:

- The columns headed 'DIF' contain the difference between the estimates of item difficulty for girls and boys;
- The columns headed '|DIF|>0.3' provide an indicator of the magnitude of the difference. The value +1 indicate that the item is easier for males than for females and the value –1 indicate that it is easier for females than for males.

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

Table 17.3
Gender DIF table for embedded attitude items[1]

Item	DIF	|DIF|>0.3	ItemId	DIF	|DIF|>0.3	ItemId	DIF	|DIF|>0.3
S408QNA	0.01	0	S485QNB	0.2	0	S425QSA	–0.18	0
S408QNB	–0.17	0	S485QNC	0.06	0	S425QSB	–0.08	0
S408QNC	–0.14	0	S498QNA	0.03	0	S425QSC	0.02	0
S413QNA	**0.39**	**1**	S498QNB	**–0.31**	**–1**	S426QSA	0.08	0
S413QNB	**0.37**	**1**	S498QNC	–0.19	0	S426QSB	0.01	0
S413QNC	0.21	0	S508QNA	0	0	S426QSC	–0.05	0
S416QNA	–0.18	0	S508QNB	–0.15	0	S438QSA	–0.04	0
S416QNB	**–0.37**	**–1**	S508QNC	–0.13	0	S438QSB	0.16	0
S428QNA	–0.19	0	S514QNA	**0.43**	**1**	S438QSC	–0.09	0
S428QNB	**–0.33**	**–1**	S514QNB	0.5	**1**	S456QSA	–0.12	0
S428QNC	–0.15	0	S514QNC	0.32	**1**	S456QSB	–0.09	0
S437QNA	0.28	0	S519QNA	–0.04	0	S456QSC	–0.02	0
S437QNB	0.26	0	S519QNB	0.22	0	S465QSA	0.20	0
S437QNC	**0.32**	**1**	S519QNC	**0.44**	**1**	S465QSB	–0.01	0
S438QNA	0.02	0	S521QNA	**0.36**	**1**	S476QSA	0.09	0
S438QNB	0.02	0	S521QNB	0.17	0	S476QSB	–0.22	0
S438QNC	–0.13	0	S524QNA	0.17	0	S476QSC	–0.27	0
S456QNA	–0.18	0	S524QNB	–0.07	0	S477QSA	0.05	0
S456QNB	–0.09	0	S524QNC	0.09	0	S477QSB	0.05	0
S456QNC	–0.04	0	S527QNA	0.13	0	S477QSC	–0.01	0
S466QNA	–0.01	0	S527QNB	0.15	0	S485QSB	0.14	0
S466QNB	0.10	0	S527QNC	–0.12	0	S485QSC	–0.05	0
S466QNC	**–0.36**	**–1**	S408QSA	0.04	0	S498QSA	0.00	0
S476QNA	**–0.34**	**–1**	S408QSB	–0.01	0	S498QSB	–0.13	0
S476QNB	–0.25	0	S408QSC	–0.06	0	S519QSA	0.02	0
S476QNC	**–0.41**	**–1**	S416QSA	0.09	0	S519QSB	0.13	0
S478QNA	–0.20	0	S416QSB	0.08	0	S519QSC	–0.04	0
S478QNB	**–0.39**	**–1**	S416QSC	–0.02	0	S527QSB	0.17	0
S478QNC	–0.23	0	S421QSA	0.27	0	S527QSC	–0.11	0
S485QNA	–0.06	0	S421QSC	0	0			

1. Absolute values greater than 0.3 are displayed in bold in this table.

While this analysis has shown the existence of DIF for some *embedded interest* items, no substantial DIF were detected for *embedded support* items.

Summary of scalability

In summary the basic psychometric characteristics of the embedded scales appear to be sound. The existence of two factors is confirmed and both the fit to the scaling model and the reliabilities of scales appear to be adequate.

The review of differential item functioning with respect to country (item-by-country interactions) and gender shows that the embedded attitude items have fewer instances of DIF (by country and gender) than do the PISA cognitive items.

RELATIONSHIP AND COMPARISONS WITH OTHER VARIABLES

Having confirmed the adequacy of the psychometric properties of the embedded attitude scales we now consider the so-called nomothetic span of these scales. Loosely speaking, nomothetic span considers the extent to which a construct relates with other constructs in expected ways. We do this by examining the relationships of the embedded attitude scales with proficiencies, student background variables and the other PISA affective scales.

Within-country student level correlations with achievement and selected background variables

Table 0.4 shows the estimated within-country student-level correlations between *embedded interest* and *embedded support* scales and *reading, mathematics* and *science* performance and highest occupational status of parents (*HISEI*). The estimates reported in Table 17.4 were computed from weighted likelihood estimates of proficiency and then disattenuated by dividing the uncorrected correlation by the square root of the product of the reliabilities for each scale.

The correlations of *embedded support* with *reading, mathematics* and *science* have medians of 0.30, 0.24 and 0.28, respectively. For reading and science, approximately 50% of the values lie between 0.25 and 0.35, whereas for mathematics the values are typically a little lower, with 50% ranging between 0.18 and 0.28.

Table 17.4
Correlation amongst attitudinal scales, performance scales and HISEI[1]

		Correlation Embedded Support (WLE) with					Correlation Embedded Interest (WLE) with			
		Science (WLE)	Maths (WLE)	Read (WLE)	Emb. Int. (WLE)	HISEI	Science (WLE)	Maths (WLE)	Read (WLE)	HISEI
OECD	Australia	**0.38**	**0.30**	**0.33**	**0.54**	0.14	0.12	0.10	0.07	0.03
	Austria	**0.28**	**0.24**	**0.29**	**0.55**	0.13	0.06	0.01	0.02	–0.02
	Belgium	**0.29**	**0.16**	**0.22**	**0.55**	0.14	0.04	–0.04	–0.05	0.01
	Canada	**0.33**	**0.27**	**0.30**	**0.57**	0.12	0.13	0.12	0.10	0.03
	Czech Republic	**0.22**	0.13	**0.20**	**0.52**	0.03	0.04	–0.01	0.04	–0.03
	Denmark	**0.34**	**0.27**	**0.30**	**0.51**	0.15	**0.16**	0.13	0.14	0.07
	Finland	**0.31**	**0.21**	**0.35**	**0.56**	0.12	0.20	0.12	**0.21**	0.06
	France	**0.33**	**0.27**	**0.28**	**0.62**	0.14	0.13	0.11	0.10	0.04
	Germany	**0.32**	**0.30**	**0.37**	**0.61**	0.11	0.09	0.07	0.03	0.01
	Greece	**0.36**	**0.27**	**0.33**	**0.59**	**0.16**	0.07	0.03	0.09	0.00
	Hungary	**0.25**	**0.18**	**0.23**	**0.58**	0.08	–0.01	–0.02	–0.01	–0.01
	Iceland	**0.42**	**0.32**	**0.37**	**0.63**	0.13	**0.28**	0.20	**0.19**	0.09
	Ireland	**0.38**	**0.30**	**0.31**	**0.57**	0.13	0.14	0.10	0.06	0.05
	Italy	**0.27**	**0.16**	**0.27**	**0.65**	0.12	–0.04	–0.06	–0.03	–0.01
	Japan	**0.26**	**0.21**	**0.21**	**0.69**	0.07	**0.19**	0.16	0.15	0.05
	Luxembourg	**0.34**	**0.26**	**0.30**	**0.57**	0.17	0.08	0.07	0.05	0.03
	Mexico	**0.24**	**0.22**	**0.26**	**0.57**	0.10	–0.05	–0.03	–0.02	–0.11
	Netherlands	**0.26**	**0.18**	**0.22**	**0.50**	0.13	0.06	0.02	–0.04	0.03
	New Zealand	**0.35**	**0.28**	**0.34**	**0.57**	0.15	0.08	0.05	0.04	0.03
	Norway	**0.42**	**0.34**	**0.37**	**0.63**	0.15	**0.27**	**0.21**	**0.21**	0.05
	Poland	**0.30**	**0.23**	**0.29**	**0.47**	0.10	–0.06	–0.08	–0.11	–0.06
	Portugal	**0.28**	**0.24**	**0.24**	**0.60**	0.11	–0.08	–0.07	–0.11	–0.11
	Korea	**0.33**	**0.30**	**0.28**	**0.55**	0.06	0.20	0.19	**0.16**	0.07
	Scotland	**0.43**	**0.34**	**0.34**	**0.53**	**0.17**	0.20	0.16	0.15	0.11
	Slovak Republic	**0.28**	**0.24**	**0.23**	**0.55**	0.11	0.01	0.01	0.01	–0.01
	Spain	**0.26**	**0.20**	**0.20**	**0.53**	0.06	0.01	–0.01	–0.07	–0.05
	Sweden	**0.38**	**0.31**	**0.43**	**0.56**	**0.19**	0.20	0.16	0.22	0.09
	Switzerland	**0.26**	**0.18**	**0.24**	**0.55**	0.11	0.12	0.07	0.08	0.04
	Turkey	**0.38**	**0.22**	**0.37**	**0.55**	0.07	0.01	–0.06	–0.02	–0.07
	United States	**0.31**	**0.28**	**0.31**	**0.53**	0.09	–0.03	–0.07	–0.06	–0.07
Partners	Azerbaijan	0.13	0.14	0.12	**0.78**	–0.02	0.01	0.08	–0.03	–0.08
	Brazil	**0.24**	0.15	**0.25**	**0.63**	0.08	–0.16	–0.18	–0.11	–0.14
	Colombia	**0.24**	**0.16**	0.15	**0.55**	0.11	0.04	0.08	0.05	–0.01
	Croatia	**0.22**	**0.18**	**0.23**	**0.51**	0.03	–0.05	–0.08	–0.06	–0.10
	Estonia	**0.26**	**0.21**	**0.24**	**0.53**	0.09	0.01	0.00	0.00	–0.02
	Hong Kong-China	**0.35**	**0.24**	**0.26**	**0.62**	0.03	**0.23**	0.18	0.14	0.02
	Israel	**0.27**	0.15	**0.22**	**0.63**	0.00	–0.01	–0.07	–0.06	–0.12
	Jordan	**0.37**	**0.31**	**0.30**	**0.76**	0.12	0.11	0.06	0.10	–0.03
	Kyrgyzstan	0.15	0.12	**0.26**	**0.82**	0.03	–0.04	0.02	0.03	0.01
	Latvia	**0.27**	**0.20**	**0.22**	**0.51**	0.09	–0.07	–0.10	–0.10	–0.05
	Lithuania	**0.32**	**0.25**	**0.29**	**0.53**	0.09	–0.04	–0.04	–0.03	–0.05
	Macao-China	**0.33**	**0.24**	**0.31**	**0.56**	0.04	0.18	0.15	0.13	0.02
	Montenegro	**0.28**	**0.17**	**0.28**	**0.51**	0.09	0.16	0.09	0.15	0.00
	Qatar	**0.41**	**0.27**	**0.40**	**0.89**	–0.01	0.05	0.00	0.08	–0.09
	Romania	**0.33**	**0.31**	**0.35**	**0.70**	0.16	–0.04	0.02	0.07	–0.08
	Russian Federation	**0.30**	**0.24**	**0.28**	**0.53**	0.08	–0.06	–0.05	–0.08	–0.04
	Serbia	**0.22**	0.14	**0.19**	**0.51**	0.07	–0.06	–0.10	–0.06	–0.09
	Slovenia	**0.26**	0.15	**0.29**	**0.59**	0.08	0.10	0.06	0.06	–0.03
	Chinese Taipei	**0.24**	**0.17**	**0.20**	**0.55**	0.07	**0.20**	**0.17**	0.13	0.03
	Thailand	**0.25**	**0.25**	**0.31**	**0.74**	0.10	**0.18**	**0.18**	**0.20**	0.00
	Tunisia	**0.44**	**0.31**	**0.41**	**0.78**	0.13	0.13	**0.18**	0.17	–0.01

1. Absolute values greater than 0.15 are displayed in bold in this table.

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

The correlations of *embedded interest* with *reading*, *mathematics* and *science* are lower and have medians of 0.06, 0.05 and 0.05, respectively. Approximately 50% of the values lie between –0.03 and 0.15 for each of the proficiencies.

Correlations of both *embedded support* and *embedded interest* with *HISEI* are lower than the correlations with achievement variables. The median for *embedded support* is 0.10, while the median for *embedded interest* is –0.01.

To provide a frame of reference for assessing whether these results are reasonable a set of parallel correlation between relevant questionnaire variables and achievement was undertaken. The questionnaire variables that were chosen where:

- Interest in science learning: *INTSCIE*;
- Enjoyment of science: *JOYSCIE*;
- General value of science: *GENSCIE*; and
- Personal value of science: *PERSCIE*.

Table 17.5
Correlations for science scale[1]

		Correlation science (WLE) with			
		INTSCIE	JOYSCIE	GENSCIE	PERSCIE
OECD	Austria	**0.23**	**0.29**	**0.24**	0.11
	Belgium	**0.37**	**0.35**	**0.21**	**0.22**
	Canada	**0.24**	**0.34**	**0.27**	**0.27**
	Czech Republic	0.19	**0.23**	**0.25**	0.11
	Denmark	**0.22**	**0.27**	**0.24**	0.18
	Finland	0.30	0.30	**0.29**	**0.26**
	France	0.30	**0.29**	**0.23**	**0.23**
	Germany	**0.23**	**0.31**	**0.26**	0.17
	Greece	**0.23**	**0.24**	**0.23**	0.19
	Hungary	0.19	**0.24**	**0.22**	0.06
	Iceland	**0.31**	**0.42**	**0.32**	**0.32**
	Ireland	**0.31**	**0.37**	0.30	**0.31**
	Italy	0.20	**0.22**	**0.25**	0.12
	Luxembourg	0.20	**0.25**	**0.26**	0.13
	Mexico	0.05	0.06	0.14	0.03
	Netherlands	**0.23**	**0.25**	**0.29**	0.18
	Norway	0.30	**0.36**	**0.33**	**0.25**
	Poland	0.15	0.15	**0.25**	0.05
	Scotland	**0.34**	**0.37**	**0.34**	**0.31**
	Slovak Republic	0.19	0.15	**0.27**	0.04
	Spain	**0.25**	**0.33**	**0.24**	**0.24**
	Sweden	**0.29**	**0.33**	0.30	0.26
Partners	Colombia	–0.08	–0.05	0.09	–0.06
	Croatia	0.18	0.12	0.18	0.03
	Estonia	0.14	0.20	**0.28**	0.17
	Hong Kong-China	**0.28**	**0.31**	0.20	0.19
	Israel	0.15	0.23	**0.22**	0.15
	Jordan	0.15	0.16	**0.25**	0.14
	Kyrgyzstan	–0.07	–0.13	0.09	–0.1
	Latvia	0.05	0.09	**0.21**	0.07
	Lithuania	0.17	0.17	**0.24**	0.14
	Montenegro	0.13	–0.04	0.15	–0.06
	Netherlands	**0.23**	**0.25**	**0.29**	0.18
	Qatar	0.07	0.13	0.20	0.13
	Romania	0.12	0.09	**0.24**	0.05
	Russian Federation	0.08	0.11	0.15	0.03
	Serbia	0.07	–0.08	0.11	–0.08
	Slovenia	0.19	0.11	**0.25**	0.12
	Tunisia	0.14	0.12	**0.21**	0.16

1. Correlations in this table are not disattenuated for unreliability of the scales. Values greater than 0.20 are displayed in bold in this table.

The first two of the above listed variables are parallels to *embedded interest* and the second two are parallels to *embedded support*.

The estimated correlations between science proficiency and each of these four questionnaire scales are given in Table 17.4. The results reported in the table are based upon the 39 countries for which the context questionnaire data had been cleaned at the time of analysis.

The correlations of *INTSCIE*, *JOYSCIE, GENSCIE*, and *PERSCIE* with science have medians of 0.17, 0.15, 0.24 and 0.14 respectively. After accounting for the fact that these correlations have not been disattenuated for measurement error it appears that the support correlations are a little lower than the corresponding values for embedded support and the interest values are a little higher.

Relationships between embedded scales and questionnaire

Of particular interest were the relationships between the variables that quantify achievement in reading, mathematics and science, and the embedded affective variables, which were gathered using the same instruments. Similarly, of interest also were the relationships between the context questionnaire interest and support variables and the embedded affective variables, which were gathered using the different instruments but were intended to tap related constructs.

An overview of these relationships is shown in Table 17.6 which reports the results of a principal components analysis that was undertaken using the final PISA database and included all 30 OECD countries. The analysis confirms that the first component is an achievement component, the second is an interest component and the third a support component.

Table 17.6
Loadings of the achievement, interest and support variables on three varimax rotated components

	Component One	Component Two	Component Three
Science	0.956	0.054	0.081
Mathematics	0.943	0.014	0.043
Reading	0.922	0.001	0.095
Interest in science learning:	0.084	0.872	0.157
Enjoyment of science:	0.107	0.814	0.253
Embedded Interest	-0.163	0.732	0.343
General value of science	0.113	0.159	0.899
Embedded support	0.133	0.390	0.698
Personal value of science	-0.005	0.525	0.639

Table 17.7 shows the correlations, for each country, of *embedded interest* and *embedded support* with the questionnaire interest variables (*INTSCIE* and *JOYSCIE*) and questionnaire support variables (*GENSCIE*, and *PERSCIE)*. The correlations show that the embedded scales are clearly related to their parallel questionnaire scales, but they do not seem to measure exactly the same constructs.

Table 17.7
Correlation between embedded attitude scales and questionnaire attitude scales

		Correlation Interest (WLE) with		Correlation Support (WLE) with	
		INTSCIE	JOYSCIE	GENSCIE	PERSCIE
OECD	Australia	0.52	0.46	0.51	0.44
	Austria	0.49	0.46	0.45	0.34
	Belgium	0.47	0.43	0.40	0.36
	Canada	0.54	0.49	0.47	0.41
	Czech Republic	0.48	0.42	0.41	0.31
	Denmark	0.53	0.50	0.41	0.33
	Finland	0.58	0.50	0.48	0.40
	France	0.51	0.48	0.42	0.38
	Germany	0.51	0.49	0.46	0.38
	Greece	0.47	0.39	0.37	0.31
	Hungary	0.47	0.39	0.40	0.34
	Iceland	0.55	0.53	0.47	0.39
	Ireland	0.51	0.47	0.47	0.40
	Italy	0.45	0.40	0.39	0.34
	Japan	0.51	0.47	0.44	0.41
	Korea	0.45	0.41	0.43	0.36
	Luxembourg	0.49	0.46	0.45	0.35
	Mexico	0.43	0.37	0.36	0.32
	Netherlands	0.57	0.50	0.42	0.35
	New Zealand	0.55	0.51	0.49	0.44
	Norway	0.53	0.50	0.49	0.42
	Poland	0.42	0.32	0.40	0.30
	Portugal	0.41	0.38	0.46	0.42
	Slovak Republic	0.42	0.44	0.43	0.32
	Spain	0.43	0.41	0.41	0.35
	Sweden	0.56	0.50	0.51	0.42
	Switzerland	0.47	0.47	0.44	0.34
	Turkey	0.49	0.46	0.53	0.43
	United Kingdom	0.51	0.45	0.51	0.40
	United States	0.51	0.43	0.49	0.44
Partners	Azerbaijan	0.25	0.26	0.25	0.22
	Brazil	0.43	0.37	0.41	0.34
	Colombia	0.36	0.27	0.36	0.31
	Croatia	0.49	0.46	0.43	0.33
	Estonia	0.52	0.45	0.42	0.34
	Hong Kong-China	0.58	0.53	0.45	0.39
	Israel	0.52	0.47	0.38	0.39
	Jordan	0.40	0.32	0.39	0.35
	Kyrgyzstan	0.38	0.32	0.34	0.27
	Latvia	0.46	0.41	0.38	0.32
	Liechtenstein	0.54	0.41	0.49	0.41
	Lithuania	0.44	0.36	0.43	0.33
	Macao-China	0.50	0.46	0.41	0.34
	Montenegro	0.48	0.39	0.40	0.30
	Qatar	0.36	0.36	0.37	0.32
	Romania	0.32	0.34	0.40	0.30
	Russian Federation	0.47	0.36	0.33	0.24
	Serbia	0.45	0.44	0.39	0.30
	Slovenia	0.48	0.42	0.44	0.34
	Chinese Taipei	0.53	0.49	0.45	0.36
	Thailand	0.43	0.37	0.43	0.43
	Tunisia	0.40	0.31	0.39	0.36

Country level correlations with achievement and selected background variables

The results reported above have all been concerned with the overall student-level or the student-level within country. In this section we consider country-level relationships.

Table 17.8 shows the rank order correlations between the country means for the five cognitive domains, the four questionnaire attitude indices and for HISEI. Negative rank order correlations are shaded.

Table 17.8
Rank order correlation five test domains, questionnaire attitude scales and HISEI

	MATH	READ	SCIE	INT	SUP	INTSCIE	GENSCIE	JOYSCIE	PERSCIE
READ	0.94								
SCIE	0.95	0.95							
INT	–0.75	–0.80	–0.74						
SUP	–0.53	–0.58	–0.54	0.85					
INTSCIE	–0.69	–0.71	–0.68	0.86	0.73				
GENSCIE	–0.48	–0.47	–0.46	0.71	0.72	0.61			
JOYSCIE	–0.59	–0.62	–0.61	0.77	0.65	0.81	0.71		
PERSCIE	–0.62	–0.58	–0.59	0.73	0.64	0.71	0.90	0.80	
HISEI	0.40	0.39	0.38	–0.60	–0.53	–0.50	–0.46	–0.45	–0.40

Table 17.9
Intra-class correlation (rho)[1]

		Cognitive scales			Embedded scales		Questionnaire scales			
		SCIE	READ	MATH	Interest	Support	INTSCIE	JOYSCIE	GENSCIE	PERSCIE
OECD	Australia	0.16	0.17	0.15	0.02	0.03	0.06	0.09	0.07	0.06
	Austria	**0.44**	**0.43**	**0.43**	0.03	0.03	0.15	0.20	0.18	0.13
	Belgium	**0.33**	**0.33**	**0.35**	0.04	0.04	0.16	0.14	0.09	0.08
	Canada	0.16	0.14	0.15	0.03	0.04	0.09	0.10	0.09	0.09
	Czech Republic	**0.45**	**0.47**	**0.46**	0.04	0.02	0.11	0.12	0.09	0.11
	Denmark	0.12	0.09	0.09	0.03	0.04	0.13	0.12	0.08	0.07
	Finland	0.04	0.04	0.04	0.02	0.01	0.07	0.05	0.06	0.04
	France	**0.43**	**0.42**	**0.41**	0.04	0.04	0.12	0.12	0.11	0.10
	Germany	**0.47**	**0.47**	**0.52**	0.02	0.04	0.14	0.13	0.09	0.07
	Greece	**0.34**	**0.26**	**0.31**	0.01	0.05	0.06	0.08	0.07	0.06
	Hungary	**0.48**	**0.49**	**0.49**	0.03	0.03	0.10	0.11	0.09	0.11
	Iceland	0.06	0.05	0.06	0.02	0.02	0.07	0.08	0.07	0.07
	Ireland	0.14	0.12	0.14	0.02	0.02	0.16	**0.21**	0.17	0.14
	Italy	**0.42**	**0.38**	**0.39**	0.05	0.06	0.17	0.14	0.11	0.14
	Japan	**0.38**	**0.40**	**0.37**	0.03	0.04	0.06	0.08	0.08	0.06
	Luxembourg	**0.24**	**0.25**	**0.23**	0.01	0.02	0.09	0.11	0.09	0.10
	Mexico	**0.25**	**0.24**	**0.22**	0.06	0.03	0.06	0.07	0.04	0.02
	Netherlands	**0.49**	**0.47**	**0.47**	0.05	0.03	0.10	0.11	0.09	0.07
	New Zealand	0.13	0.11	0.13	0.03	0.04	0.07	0.07	0.08	0.08
	Norway	0.06	0.06	0.04	0.03	0.03	0.10	0.12	0.10	0.06
	Poland	0.16	0.16	0.10	0.03	0.01	0.09	0.09	0.08	0.08
	Portugal	**0.25**	**0.24**	**0.23**	0.05	0.03	0.20	0.18	0.15	0.18
	Korea	**0.29**	**0.30**	**0.27**	0.03	0.05	0.12	0.19	0.08	0.10
	Scotland	0.14	0.13	0.09	0.02	0.03	0.14	0.12	0.07	0.10
	Slovak Republic	**0.33**	**0.36**	**0.35**	0.04	0.03	0.17	0.12	0.12	0.14
	Spain	0.12	0.12	0.12	0.04	0.03	0.09	0.10	0.07	0.07
	Sweden	0.08	0.08	0.07	0.04	0.04	0.10	0.11	0.12	0.10
	Switzerland	**0.26**	**0.24**	**0.21**	0.03	0.04	0.12	0.14	0.11	0.10
	Turkey	**0.42**	**0.40**	**0.33**	0.04	0.05	0.09	0.11	0.12	0.10
	United States	0.20	0.19	0.18	0.05	0.02	0.05	0.06	0.06	0.05
Partners	Azerbaijan	**0.27**	0.19	**0.24**	0.09	0.05	0.11	0.13	0.13	0.14
	Brazil	**0.34**	**0.32**	**0.25**	0.04	0.02	0.11	0.14	0.10	0.12
	Colombia	0.19	**0.21**	0.13	0.04	0.01	0.11	0.11	0.06	0.10
	Croatia	**0.32**	**0.29**	**0.31**	0.03	0.02	0.06	0.05	0.05	0.04
	Estonia	0.15	0.15	0.20	0.05	0.02	0.08	0.09	0.08	0.07
	Hong Kong-China	**0.29**	**0.29**	**0.26**	0.01	0.02	0.16	0.12	0.10	0.10
	Israel	**0.24**	**0.28**	**0.27**	0.16	0.05	0.14	0.17	0.18	0.18
	Jordan	0.18	0.15	0.16	0.02	0.04	0.15	0.15	0.13	0.13
	Kyrgyzstan	**0.22**	**0.22**	**0.21**	0.05	0.02	0.09	0.10	0.09	0.07
	Latvia	0.14	0.15	0.14	0.06	0.03	0.04	0.04	0.04	0.02
	Liechtenstein	**0.37**	**0.33**	**0.42**	0.03	0.02	0.10	0.11	0.07	0.09
	Lithuania	**0.21**	**0.23**	0.20	0.06	0.03	0.11	0.12	0.12	0.05
	Macao – China	**0.21**	0.17	0.17	0.02	0.02	0.20	0.17	0.08	0.12
	Montenegro	0.19	0.16	0.20	0.03	0.01	0.09	0.09	0.10	0.09
	Qatar	**0.27**	**0.30**	**0.34**	0.06	0.10	0.07	0.06	0.06	0.07
	Romania	**0.36**	**0.36**	**0.33**	0.04	0.05	0.10	0.07	0.11	0.08
	Russian Federation	0.19	0.19	0.18	0.07	0.03	0.14	0.13	0.08	0.12
	Serbia	**0.30**	**0.29**	**0.28**	0.07	0.01	0.08	0.11	0.11	0.13
	Slovenia	**0.50**	**0.45**	**0.54**	0.04	0.04	0.19	0.18	0.17	0.17
	Chinese Taipei	**0.43**	**0.42**	**0.35**	0.04	0.03	0.07	0.08	0.05	0.04
	Thailand	**0.31**	**0.29**	**0.28**	0.04	0.03	0.12	0.09	0.08	0.08
	Tunisia	**0.28**	**0.29**	**0.23**	0.03	0.05	0.08	0.08	0.08	0.07

1. Values greater than 0.20 are displayed in bold in this table.

Rank order correlation coefficients between cognitive scales and attitude scales (both embedded and questionnaire) are negative at country level. The strongest negative relationship is between country ranks in *embedded interest* and performance. All attitude scales have strong positive rank correlations with each other. *HISEI* has a positive correlation with cognitive scales and a negative correlation with attitude scales.

Variance decomposition

Table 17.9 provides the intra-class correlation for each country and each cognitive domain; each embedded attitudinal domain and each questionnaire index. The intra-class correlation can be interpreted as the percentage of the total variance that is accounted for by differences among schools

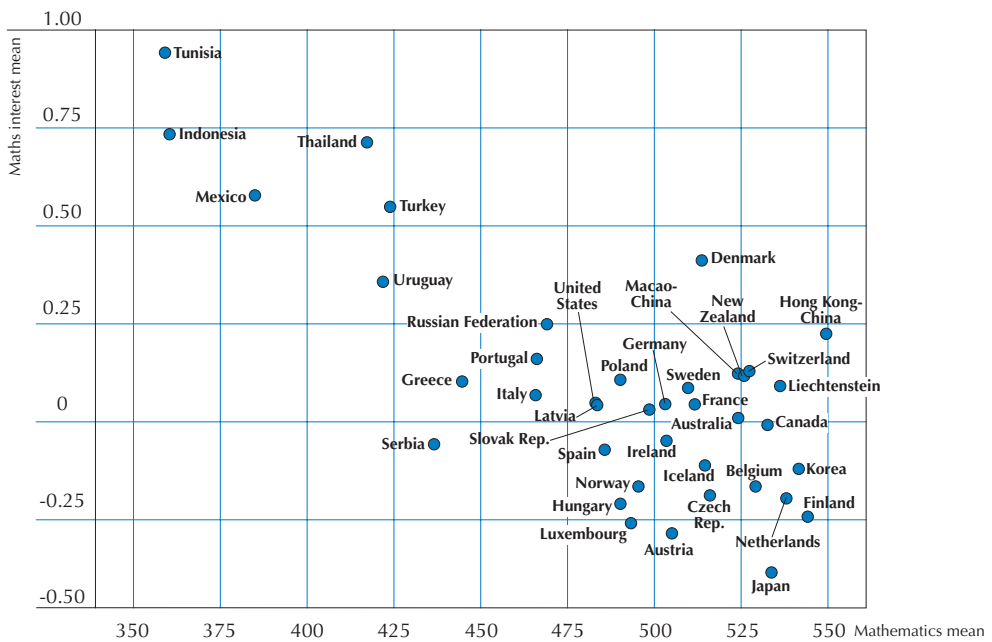
For *mathematics*, *reading* and *science* the intra-class correlation coefficient is greater than 0.20 for a number of countries, but for both the *embedded interest* and *embedded support* scale it is small for all countries. The questionnaire scales also have small intra-class correlations, although slightly larger than the embedded attitude scales. This observation is consistent with questionnaire results from previous cycles.

Observations from other cross-national data collections

We conclude the chapter by noting the relationships between similar attitudinal variables and achievement variables in PISA 2000 and 2003.

In PISA 2000 the variable closest to interest in reading (the major domain) was *JOYREAD*. For the 43 participating countries in PISA 2000 and PISA Plus the median within-country between-student correlation between reading achievement and *JOYREAD* was 0.30, with 50% of the values lying between 0.27 and 0.40. At the country level the correlation between mean reading achievement and mean *JOYREAD* was –0.63.

Figure 17.3
Scatterplot of mean mathematics interest against mean mathematics for PISA 2003



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

In PISA 2003 the interest variable *INTMATH* – mathematics being the major domain – was a close match to the INTSCIE variable included in PISA 2006. For the 40 participating countries in PISA 2003 the median within-country between-student correlation between reading achievement and *INTMATH* was 0.14, with 50% of the values lying between 0.10 and 0.24. At the country level the correlation between mean reading achievement and mean *INTMATH* was –0.76.

The country-level correlation for interest in mathematics and mathematics is shown in Figure 17.3. The correlation and the scatterplot is quite consistent with results that are observed for the attitude scales in PISA 2006 – both for the embedded attitude scales and the attitude scales that are included in the context questionnaires. Furthermore, the results are consistent with those found in other international studies such as the Trends in Mathematics and Science Study (*TIMSS*).

Summary of relations with other variables

The embedded items behave in expected and predictable ways with the other PISA variables. Principal component analysis supports that they are distinct dimensions that correlate appropriately with parallel scales that were included in the context questionnaires. Further their correlations, both at the student-level within-country and at the country level, with various other variables are consistent with observations that are made in other PISA data collections and in other studies.

CONCLUSION

The purpose of this chapter was to present analyses that support the use of the embedded scales as constructs that have the potential to provide useful and valid across-country comparisons. The purpose was not to present a comprehensive set of analyses that fully explore the relations between the embedded attitude scales and other PISA variables – such analyses will be reported elsewhere in research that draws upon the PISA databases.

The main conclusions are that embedded scales have been well constructed and are strongly supported by theory that is articulated in the PISA 2006 assessment frameworks (OECD, 2006). Statistical analysis indicates that from a psychometric perspective the embedded scales are equivalent, in terms of robustness and cross-participant validity, to the PISA cognitive scales.

In terms of their basic relationships with other variables, the embedded items generally behave in ways that are consistent with other affective variables. Our discussion of this, however, does suggest a number of important research issues that need to be explored with PISA and other data sources. Some issues that would seem worthy of pursuing are:

Why do affective variables (both embedded and otherwise) typically show a much lower intra-class correlation than do achievement variables, and to a lesser extent than do other student contextual variables?

Why do so many affective variables (both embedded and otherwise) have a negative correlation at the country level with performance measures? To what extent are these negative correlations simply examples of ecological fallacies, interpretable and important findings or cultural and misleading artifacts in the response behaviours of students?

Is there anything to be learned from the fact that lower correlations are observed between the embedded interest scales and student proficiency than between the questionnaire interest scales and student proficiency?

Notes

1. For an elaboration of these scales, see the *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006* (OECD, 2006).

2. The reliability results presented in Chapter 12 and the IRT item parameter estimates reported in Appendix 7 are based upon a properly weighted and cleaned calibration sample and may differ a little from those reported here.

3. UH booklet is an optional one hour in length booklet, which some countries implemented in special educational needs settings (see Chapter 3).

4. Magnitudes and directions of booklet one factor loadings are very similar for the continuous and categorical approaches.

5. Note that these figures differ from those reported in Chapter 13 because the values reported in Chapter 13 were estimated using the final database.

Reader's Guide

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

OECD countries

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

Partner countries and economies

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table of contents

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

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

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

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

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

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

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

LIST OF BOXES

LIST OF FIGURES

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

LIST OF TABLES

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

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

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

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

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

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