



# Validation of the Embedded Attitudinal Scales

Introduction.....	352
International scalability.....	353
▪ Analysis of item dimensionality with exploratory and confirmatory factor analysis.....	353
▪ Fit to item response model.....	353
▪ Reliability.....	355
▪ Differential item functioning.....	355
▪ Summary of scalability.....	357
Relationship and comparisons with other variables.....	357
▪ Within-country student level correlations with achievement and selected background variables.....	358
▪ Relationships between embedded scales and questionnaire.....	360
▪ Country level correlations with achievement and selected background variables.....	361
▪ Variance decomposition.....	363
▪ Observations from other cross-national data collections.....	363
▪ Summary of relations with other variables.....	364
Conclusion.....	364



## 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*.<sup>1</sup> 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.<sup>2</sup> 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.<sup>3</sup>

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*<sup>®</sup> (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.<sup>4</sup>

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*<sup>®</sup> 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

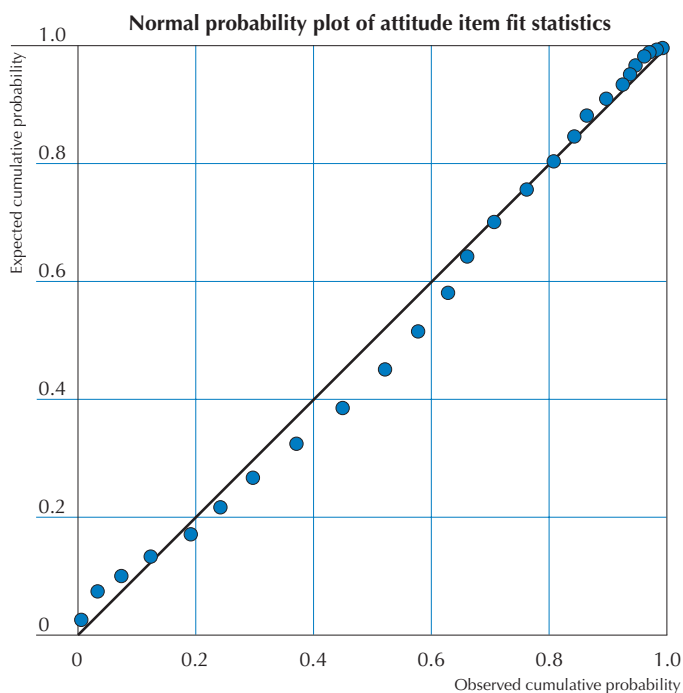


Table 17.1 shows that the estimated latent correlation between *embedded interest* and *embedded support* is 0.623. This is very similar to the corresponding value, 0.594, that was estimated using the CFA.

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  $-0.15$ .<sup>5</sup> 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
<i>For interest in...</i>						
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
<i>For support in...</i>						
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.

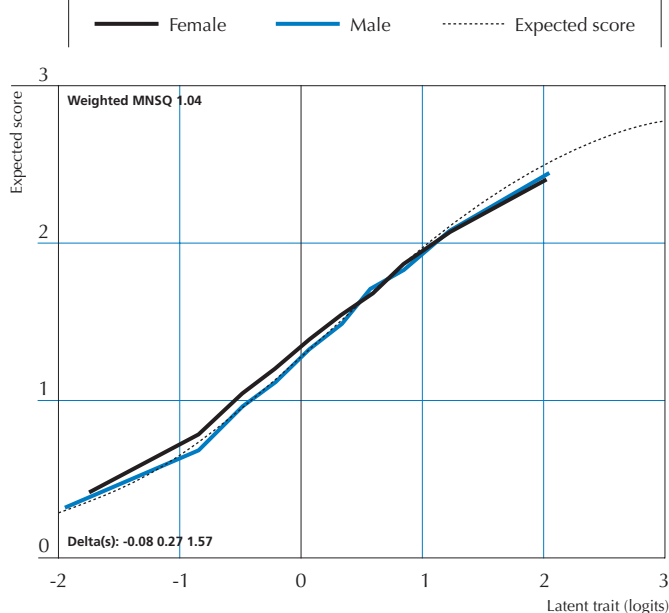


Item-by-country interactions indicate the degree of parameter invariance across countries. The results show that parameters for items measuring both *embedded interest* and *embedded support* tend to be fairly stable across countries. Furthermore, there were no items with a discrimination value less 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.



**Table 17.3**  
Gender DIF table for embedded attitude items<sup>1</sup>

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	<b>0.39</b>	<b>1</b>	S498QNB	<b>-0.31</b>	<b>-1</b>	S426QSA	0.08	0
S413QNB	<b>0.37</b>	<b>1</b>	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	<b>-0.37</b>	<b>-1</b>	S508QNC	-0.13	0	S438QSB	0.16	0
S428QNA	-0.19	0	S514QNA	<b>0.43</b>	<b>1</b>	S438QSC	-0.09	0
S428QNB	<b>-0.33</b>	<b>-1</b>	S514QNB	0.5	<b>1</b>	S456QSA	-0.12	0
S428QNC	-0.15	0	S514QNC	0.32	<b>1</b>	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	<b>0.32</b>	<b>1</b>	S519QNC	<b>0.44</b>	<b>1</b>	S465QSB	-0.01	0
S438QNA	0.02	0	S521QNA	<b>0.36</b>	<b>1</b>	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	<b>-0.36</b>	<b>-1</b>	S408QSA	0.04	0	S498QSA	0.00	0
S476QNA	<b>-0.34</b>	<b>-1</b>	S408QSB	-0.01	0	S498QSB	-0.13	0
S476QNB	-0.25	0	S408QSC	-0.06	0	S519QSA	0.02	0
S476QNC	<b>-0.41</b>	<b>-1</b>	S416QSA	0.09	0	S519QSB	0.13	0
S478QNA	-0.20	0	S416QSB	0.08	0	S519QSC	-0.04	0
S478QNB	<b>-0.39</b>	<b>-1</b>	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<sup>1</sup>

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

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





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<sup>1</sup>

		Correlation science (WLE) with			
		INTSCIE	JOYSCIE	GENSCIE	PERSCIE
<i>OECD</i>	Austria	<b>0.23</b>	<b>0.29</b>	<b>0.24</b>	0.11
	Belgium	<b>0.37</b>	<b>0.35</b>	<b>0.21</b>	<b>0.22</b>
	Canada	<b>0.24</b>	<b>0.34</b>	<b>0.27</b>	<b>0.27</b>
	Czech Republic	0.19	<b>0.23</b>	<b>0.25</b>	0.11
	Denmark	<b>0.22</b>	<b>0.27</b>	<b>0.24</b>	0.18
	Finland	0.30	0.30	<b>0.29</b>	<b>0.26</b>
	France	0.30	<b>0.29</b>	<b>0.23</b>	<b>0.23</b>
	Germany	<b>0.23</b>	<b>0.31</b>	<b>0.26</b>	0.17
	Greece	<b>0.23</b>	<b>0.24</b>	<b>0.23</b>	0.19
	Hungary	0.19	<b>0.24</b>	<b>0.22</b>	0.06
	Iceland	<b>0.31</b>	<b>0.42</b>	<b>0.32</b>	<b>0.32</b>
	Ireland	<b>0.31</b>	<b>0.37</b>	0.30	<b>0.31</b>
	Italy	0.20	<b>0.22</b>	<b>0.25</b>	0.12
	Luxembourg	0.20	<b>0.25</b>	<b>0.26</b>	0.13
	Mexico	0.05	0.06	0.14	0.03
	Netherlands	<b>0.23</b>	<b>0.25</b>	<b>0.29</b>	0.18
	Norway	0.30	<b>0.36</b>	<b>0.33</b>	<b>0.25</b>
	Poland	0.15	0.15	<b>0.25</b>	0.05
	Scotland	<b>0.34</b>	<b>0.37</b>	<b>0.34</b>	<b>0.31</b>
	Slovak Republic	0.19	0.15	<b>0.27</b>	0.04
Spain	<b>0.25</b>	<b>0.33</b>	<b>0.24</b>	<b>0.24</b>	
Sweden	<b>0.29</b>	<b>0.33</b>	0.30	0.26	
<i>Partners</i>	Colombia	-0.08	-0.05	0.09	-0.06
	Croatia	0.18	0.12	0.18	0.03
	Estonia	0.14	0.20	<b>0.28</b>	0.17
	Hong Kong-China	<b>0.28</b>	<b>0.31</b>	0.20	0.19
	Israel	0.15	0.23	<b>0.22</b>	0.15
	Jordan	0.15	0.16	<b>0.25</b>	0.14
	Kyrgyzstan	-0.07	-0.13	0.09	-0.1
	Latvia	0.05	0.09	<b>0.21</b>	0.07
	Lithuania	0.17	0.17	<b>0.24</b>	0.14
	Montenegro	0.13	-0.04	0.15	-0.06
	Netherlands	<b>0.23</b>	<b>0.25</b>	<b>0.29</b>	0.18
	Qatar	0.07	0.13	0.20	0.13
	Romania	0.12	0.09	<b>0.24</b>	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	<b>0.25</b>	0.12
	Tunisia	0.14	0.12	<b>0.21</b>	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	
<b>OECD</b>	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	
<b>Partners</b>	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 ( $\rho$ )<sup>1</sup>

	Cognitive scales			Embedded scales		Questionnaire scales			
	SCIE	READ	MATH	Interest	Support	INTSCIE	JOYSCIE	GENSCIE	PERSCIE
<b>OECD</b>									
Australia	0.16	0.17	0.15	0.02	0.03	0.06	0.09	0.07	0.06
Austria	<b>0.44</b>	<b>0.43</b>	<b>0.43</b>	0.03	0.03	0.15	0.20	0.18	0.13
Belgium	<b>0.33</b>	<b>0.33</b>	<b>0.35</b>	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	<b>0.45</b>	<b>0.47</b>	<b>0.46</b>	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	<b>0.43</b>	<b>0.42</b>	<b>0.41</b>	0.04	0.04	0.12	0.12	0.11	0.10
Germany	<b>0.47</b>	<b>0.47</b>	<b>0.52</b>	0.02	0.04	0.14	0.13	0.09	0.07
Greece	<b>0.34</b>	<b>0.26</b>	<b>0.31</b>	0.01	0.05	0.06	0.08	0.07	0.06
Hungary	<b>0.48</b>	<b>0.49</b>	<b>0.49</b>	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	<b>0.21</b>	0.17	0.14
Italy	<b>0.42</b>	<b>0.38</b>	<b>0.39</b>	0.05	0.06	0.17	0.14	0.11	0.14
Japan	<b>0.38</b>	<b>0.40</b>	<b>0.37</b>	0.03	0.04	0.06	0.08	0.08	0.06
Luxembourg	<b>0.24</b>	<b>0.25</b>	<b>0.23</b>	0.01	0.02	0.09	0.11	0.09	0.10
Mexico	<b>0.25</b>	<b>0.24</b>	<b>0.22</b>	0.06	0.03	0.06	0.07	0.04	0.02
Netherlands	<b>0.49</b>	<b>0.47</b>	<b>0.47</b>	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	<b>0.25</b>	<b>0.24</b>	<b>0.23</b>	0.05	0.03	0.20	0.18	0.15	0.18
Korea	<b>0.29</b>	<b>0.30</b>	<b>0.27</b>	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	<b>0.33</b>	<b>0.36</b>	<b>0.35</b>	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	<b>0.26</b>	<b>0.24</b>	<b>0.21</b>	0.03	0.04	0.12	0.14	0.11	0.10
Turkey	<b>0.42</b>	<b>0.40</b>	<b>0.33</b>	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
<b>Partners</b>									
Azerbaijan	<b>0.27</b>	0.19	<b>0.24</b>	0.09	0.05	0.11	0.13	0.13	0.14
Brazil	<b>0.34</b>	<b>0.32</b>	<b>0.25</b>	0.04	0.02	0.11	0.14	0.10	0.12
Colombia	0.19	<b>0.21</b>	0.13	0.04	0.01	0.11	0.11	0.06	0.10
Croatia	<b>0.32</b>	<b>0.29</b>	<b>0.31</b>	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	<b>0.29</b>	<b>0.29</b>	<b>0.26</b>	0.01	0.02	0.16	0.12	0.10	0.10
Israel	<b>0.24</b>	<b>0.28</b>	<b>0.27</b>	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	<b>0.22</b>	<b>0.22</b>	<b>0.21</b>	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	<b>0.37</b>	<b>0.33</b>	<b>0.42</b>	0.03	0.02	0.10	0.11	0.07	0.09
Lithuania	<b>0.21</b>	<b>0.23</b>	0.20	0.06	0.03	0.11	0.12	0.12	0.05
Macao - China	<b>0.21</b>	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	<b>0.27</b>	<b>0.30</b>	<b>0.34</b>	0.06	0.10	0.07	0.06	0.06	0.07
Romania	<b>0.36</b>	<b>0.36</b>	<b>0.33</b>	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	<b>0.30</b>	<b>0.29</b>	<b>0.28</b>	0.07	0.01	0.08	0.11	0.11	0.13
Slovenia	<b>0.50</b>	<b>0.45</b>	<b>0.54</b>	0.04	0.04	0.19	0.18	0.17	0.17
Chinese Taipei	<b>0.43</b>	<b>0.42</b>	<b>0.35</b>	0.04	0.03	0.07	0.08	0.05	0.04
Thailand	<b>0.31</b>	<b>0.29</b>	<b>0.28</b>	0.04	0.03	0.12	0.09	0.08	0.08
Tunisia	<b>0.28</b>	<b>0.29</b>	<b>0.23</b>	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.

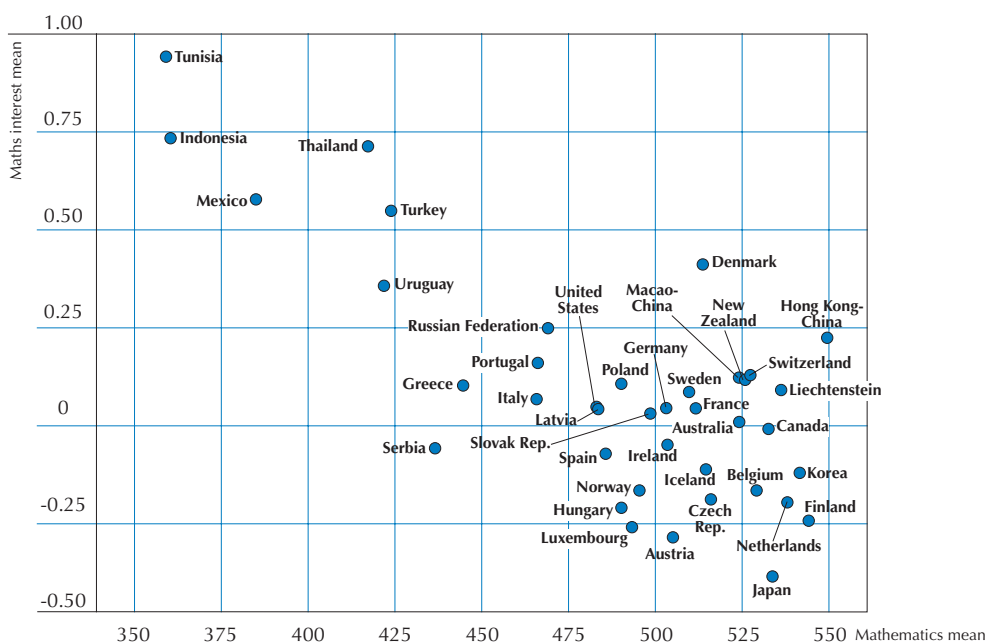
### Observations from other cross-national data collections

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





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	Kyrgyzstan
LIE	Liechtenstein
LTU	Lithuania
LVA	Latvia
LVL	Latvia (Latvian Community)
LVR	Latvia (Russian Community)
MAC	Macao-China
MNE	Montenegro
QAT	Qatar
ROU	Romania
RUS	Russian Federation
SRB	Serbia
SVN	Slovenia
TAP	Chinese Taipei
THA	Thailand
TUN	Tunisia
URY	Uruguay



# References

- Adams, R.J., Wilson, M. & Wang, W.C.** (1997), The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, No. 21, pp. 1-23.
- Adams, R.J., Wilson, M. R. & Wu, M.L.** (1997), Multilevel item response models: An approach to errors in variables regression, *Journal of Educational and Behavioural Statistics*, No. 22 (1), pp. 46-75.
- Adams, R.J. & Wu, M.L.** (2002), *PISA 2000 Technical Report*, OECD, Paris.
- Bollen, K.A. & Long, S.J.** (1993) (eds.), *Testing Structural Equation Models*, Newbury Park: London.
- Beaton, A.E.** (1987), Implementing the new design: The NAEP 1983-84 technical report (Rep. No. 15-TR-20), Princeton, NJ: Educational Testing Service.
- Buchmann, C.** (2000), Family structure, parental perceptions and child labor in Kenya: What factors determine who is enrolled in school? *Soc. Forces*, No. 78, pp. 1349-79.
- Buchmann, C.** (2002), Measuring Family Background in International Studies of Education: Conceptual Issues and Methodological Challenges, in Porter, A.C. and Gamoran, A. (eds.). *Methodological Advances in Cross-National Surveys of Educational Achievement* (pp. 150-97), Washington, DC: National Academy Press.
- Creemers, B.P.M.** (1994), *The Effective Classroom*, London: Cassell.
- Cochran, W.G.** (1977), *Sampling techniques*, third edition, New York, NY: John Wiley and Sons.
- Ganzeboom, H.B.G., de Graaf, P.M. & Treiman, D.J.** (1992), A standard international socio-economic index of occupational status, *Social Science Research*, No. 21, pp. 1-56.
- Ganzeboom H.B. & Treiman, D.J.** (1996), Internationally comparable measures of occupational status for the 1988 international standard classification of occupations, *Social Science Research*, No. 25, pp. 201-239.
- Grisay, A.** (2003), Translation procedures in OECD/PISA 2000 international assessment, *Language Testing*, No. 20 (2), pp. 225-240.
- Hambleton, R.K., Swaminathan, H. & Rogers, H.J.** (1991), *Fundamentals of item response theory*, Newbury Park, London, New Delhi: SAGE Publications.
- Hambleton, R.K., Merenda, P.F. & Spielberger, C.D.** (2005), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*, IEA Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.
- Harkness, J.A., Van de Vijver, F.J.R. & Mohler, P.Ph** (2003), *Cross-Cultural Survey Methods*, Wiley-Interscience, John Wiley & Sons, Inc., Hoboken, New Jersey.
- Harvey-Beavis, A.** (2002), Student and School Questionnaire Development, in R.J. Adams and M.L. Wu (eds.), *PISA 2000 Technical Report*, (pp. 33-38), OECD, Paris.
- International Labour Organisation (ILO)** (1990), *International Standard Classification of Occupations: ISCO-88*. Geneva: International Labour Office.
- Jöreskog, K.G. & Sörbom, Dag** (1993), *LISREL 8 User's Reference Guide*, Chicago: SSI.
- Judkins, D.R.** (1990), Fay's Method of Variance Estimation, *Journal of Official Statistics*, No. 6 (3), pp. 223-239.
- Kaplan, D.** (2000), *Structural equation modeling: Foundation and extensions*, Thousand Oaks: SAGE Publications.
- Keyfitz, N.** (1951), Sampling with probabilities proportionate to science: Adjustment for changes in probabilities, *Journal of the American Statistical Association*, No. 46, American Statistical Association, Alexandria, pp. 105-109.
- Kish, L.** (1992), Weighting for Unequal, *Pi. Journal of Official Statistics*, No. 8 (2), pp. 183-200.
- LISREL** (1993), K.G. Jöreskog & D. Sörbom, [computer software], Lincolnwood, IL: Scientific Software International, Inc.
- Lohr, S.L.** (1999), *Sampling: Design and Analysis*, Duxberry: Pacific Grove.
- Macaskill, G., Adams, R.J. & Wu, M.L.** (1998), Scaling methodology and procedures for the mathematics and science literacy, advanced mathematics and physics scale, in M. Martin and D.L. Kelly, Editors, *Third International Mathematics and Science Study, technical report Volume 3: Implementation and analysis*, Boston College, Chestnut Hill, MA.
- Masters, G.N. & Wright, B.D.** (1997), The Partial Credit Model, in W.J. van der Linden, & R.K. Hambleton (eds.), *Handbook of Modern Item Response Theory* (pp. 101-122), New York/Berlin/Heidelberg: Springer.

- Mislevy, R.J.** (1991), Randomization-based inference about latent variables from complex samples, *Psychometrika*, No. 56, pp. 177-196.
- Mislevy, R.J., Beaton, A., Kaplan, B.A. & Sheehan, K.** (1992), Estimating population characteristics from sparse matrix samples of item responses, *Journal of Educational Measurement*, No. 29 (2), pp. 133-161.
- Mislevy, R.J. & Sheehan, K.M.** (1987), Marginal estimation procedures, in Beaton, A.E., Editor, 1987. *The NAEP 1983-84 technical report*, National Assessment of Educational Progress, Educational Testing Service, Princeton, pp. 293-360.
- Mislevy, R.J. & Sheehan, K.M.** (1989), Information matrices in latent-variable models, *Journal of Educational Statistics*, No. 14, pp. 335-350.
- Mislevy, R.J. & Sheehan, K.M.** (1989), The role of collateral information about examinees in item parameter estimation, *Psychometrika*, No. 54, pp. 661-679.
- Monseur, C. & Berezner, A.** (2007), The Computation of Equating Errors in International Surveys in Education, *Journal of Applied Measurement*, No. 8 (3), 2007, pp. 323-335.
- Monseur, C.** (2005), An exploratory alternative approach for student non response weight adjustment, *Studies in Educational Evaluation*, No. 31 (2-3), pp. 129-144.
- Muthen, B. & L. Muthen** (1998), [computer software], *Mplus* Los Angeles, CA: Muthen & Muthen.
- Muthen, B., du Toit, S.H.C. & Spisic, D.** (1997), *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*, unpublished manuscript.
- OECD** (1999), *Classifying Educational Programmes. Manual for ISCED-97 Implementation in OECD Countries*, OECD, Paris.
- OECD** (2003), *Literacy Skills for the World of Tomorrow: Further results from PISA 2000*, OECD, Paris.
- OECD** (2004), *Learning for Tomorrow's World – First Results from PISA 2003*, OECD, Paris.
- OECD** (2005), *Technical Report for the OECD Programme for International Student Assessment 2003*, OECD, Paris.
- OECD** (2006), *Assessing Scientific, Reading and Mathematical Literacy: A framework for PISA 2006*, OECD, Paris.
- OECD** (2007), *PISA 2006: Science Competencies for Tomorrow's World*, OECD, Paris.
- PISA Consortium** (2006), *PISA 2006 Main Study Data Management Manual*, [https://mypisa.acer.edu.au/images/mypisadoc/opmanual/pisa2006\\_data\\_management\\_manual.pdf](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/area06routitsky\\_berezner.pdf](https://mypisa.acer.edu.au/images/mypisadoc/area06routitsky_berezner.pdf)
- Rust, K.** (1985), Variance Estimation for Complex Estimators in Sample Surveys, *Journal of Official Statistics*, No. 1, pp. 381-397.
- Rust, K.F. & Rao, J.N.K.** (1996), Variance Estimation for Complex Surveys Using Replication Techniques, *Survey Methods in Medical Research*, No. 5, pp. 283-310.
- Shao, J.** (1996), Resampling Methods in Sample Surveys (with Discussion), *Statistics*, No. 27, pp. 203-254.
- Särndal, C.-E., Swensson, B. & Wretman, J.** (1992), *Model Assisted Survey Sampling*, New York: Springer-Verlag.
- SAS® CALIS** (1992), W. Hartmann [computer software], Cary, NC: SAS Institute Inc.
- Scheerens, J.** (1990), School effectiveness and the development of process indicators of school functioning, *School effectiveness and school improvement*, No. 1, pp. 61-80.
- Scheerens, J. & Bosker, R.J.** (1997), *The Foundations of School Effectiveness*, Oxford: Pergamon.
- Schulz, W.** (2002), Constructing and Validating the Questionnaire composites, in R.J. Adams and M.L. Wu (eds.), *PISA 2000 Technical Report*, OECD, Paris.
- Schulz, W.** (2004), Mapping Student Scores to Item Responses, in W. Schulz and H. Sibberns (eds.), *IEA Civic Education Study, Technical Report* (pp. 127-132), Amsterdam: IEA.
- Schulz, W.** (2006a), *Testing Parameter Invariance for Questionnaire Indices using Confirmatory Factor Analysis and Item Response Theory*, Paper presented at the Annual Meetings of the American Educational Research Association (AERA) in San Francisco, 7-11 April.
- Schulz, W.** (2006b), *Measuring the socio-economic background of students and its effect on achievement in PISA 2000 and PISA 2003*, Paper presented at the Annual Meetings of the American Educational Research Association (AERA) in San Francisco, 7-11 April.
- Thorndike, R.L.** (1973), *Reading comprehension in fifteen countries*, New York, Wiley: and Stockholm: Almqvist & Wiksell.
- Travers, K.J. & Westbury, I.** (1989), *The IEA Study of Mathematics I: Analysis of Mathematics Curricula*, Oxford: Pergamon Press.



- Travers, K.J., Garden R.A. & Rosier, M.** (1989), Introduction to the Study, in Robitaille, D. A. and Garden, R. A. (eds), *The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics Curricula*, Oxford: Pergamon Press.
- Verhelst, N.** (2002), Coder and Marker Reliability Studies, in R.J. Adams & M.L. Wu (eds.), *PISA 2000 Technical Report*. OECD, Paris.
- Walberg, H.J.** (1984), Improving the productivity of American schools, *Educational Leadership*, No. 41, pp. 19-27.
- Walberg, H.** (1986), Synthesis of research on teaching, in M. Wittrock (ed.), *Handbook of research on teaching* (pp. 214-229), New York: Macmillan.
- Walker, M.** (2006), *The choice of Likert or dichotomous items to measure attitudes across culturally distinct countries in international comparative educational research*. Paper presented at the Annual Meetings of the American Educational Research Association (AERA) in San Francisco, 7-11 April.
- Walker, M.** (2007), Ameliorating Culturally-Based Extreme Response Tendencies To Attitude items, *Journal of Applied Measurement*, No. 8, pp. 267-278.
- Warm, T.A.** (1989), Weighted Likelihood Estimation of Ability in Item Response Theory, *Psychometrika*, No. 54 (3), pp. 427-450.
- Westat** (2007), *WesVar<sup>®</sup> 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<sup>®</sup>: Multi-Aspect Test Software* [computer program manual], Camberwell, Vic.: Australian Council for Educational Research.



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

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



# Table of contents

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

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



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



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



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



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



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



## LIST OF BOXES

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

## LIST OF FIGURES

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



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

## LIST OF TABLES

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

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



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



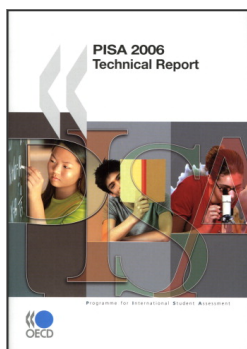


Table 12.26	Per cent correct for science link items in PISA 2000 and PISA 2003 .....	243
Table 12.27	Per cent correct for science link items in PISA 2003 and PISA 2006 .....	245
Table 12.28	Link error estimates .....	247
<hr/>		
Table 13.1	Variance components for mathematics.....	255
Table 13.2	Variance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



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

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



**From:**  
**PISA 2006 Technical Report**

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

**Please cite this chapter as:**

OECD (2009), "Validation of the Embedded Attitudinal Scales", in *PISA 2006 Technical Report*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264048096-18-en>

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

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

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