17

Questionnaire Development Issues

This chapter is concerned with a number of development issues in relation to the questionnaires used in PISA 2012. These issues include:

- the use of alternative question formats in an effort to deal with artifactual influences resulting from differences in response styles, both between respondents within countries but more so between countries;
- the lessons learned from the transition to a computer-delivered school questionnaire;
- comparative analyses undertaken because of the need to change from ISCO-88 to ISCO-08; and
- the implications of using rotated questionnaires for approaches to analysis.

ALTERNATIVE ITEM FORMATS

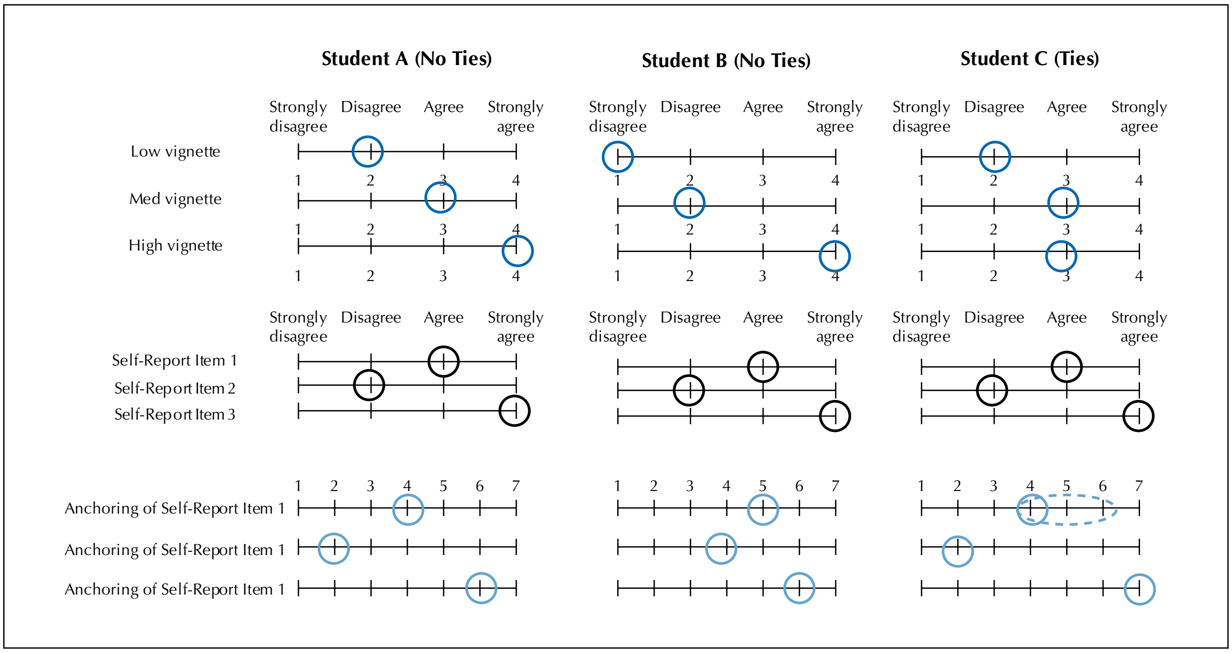
As discussed in Chapter 3, one of the major challenges of an international study such as PISA is the cross-cultural validity and applicability of all instruments. For PISA 2012, four specific alternative methods where explored in both the Field Trial and Main Survey as approaches to correcting for the impact of culturally-derived response styles that might bias the results (Kyllonen, Lietz and Roberts, 2010). In this section we report the evidence from the Field Trial and Main Survey concerning these methods. The four methods explored where anchoring vignettes, signal detection de-biasing based on the over-claiming technique, forced choice items, and situational judgement tests.

Anchoring vignettes

The original nonparametric anchoring vignette method was extended so that multiple items could be anchored based on the same set of anchoring vignettes (Bertling and Kyllonen, 2013). The scoring procedure is illustrated in the following, based on the classroom management vignettes.

As shown in Figure 17.1, the three vignettes capture three different levels of classroom management that can be described as low, medium and high (for actual wording of vignettes, see Chapter 3). Students were asked to read the vignettes and indicate their level of agreement with the statement that the described teacher is in control of his or her classroom using the same 4-point agreement scale that is also used for most questionnaire indices in the student questionnaire. Depending on their rating standards and their interpretation of the four levels of the agreement scale, students might place the three vignettes on different agreement categories. For instance, one student might "agree" that

■ Figure 17.1 ■
Illustration of scoring based on vignettes for three hypothetical students



Source: Bertling and Kyllonen (2013).

a teacher described in the first vignette is in control of his/her classroom while another student might “strongly agree” or “disagree” with this statement. Since the actual levels of teachers’ classroom management presented in the vignettes are invariant over respondents, differences in students’ response to the vignettes signal that students differ with regard to how they interpret the response scale, and that any comparisons based on raw item responses might have validity problems.

The alternative scoring based on the vignettes addresses this problem. When items were scored based on vignettes, numerical values for student responses were not assigned based on the concrete response option chosen (e.g., the value 4 for “strongly agree” and 3 for “agree”) but based on the self-report answer relative to the personal standard captured by the respondent’s individual rating of the three vignettes that form one set. Regardless of where on the 4-point agreement scale a student places the vignettes, a student’s self-report can be scored relative to his/her rating of low, medium and high for the vignettes. Based on this approach, in PISA 2012, students’ responses on the original 4-point agreement scale were re-scaled into a 7-point scale representing all possible relative rank comparisons of students’ self-reports and their rating of the vignettes. On this 7-point scale, the value 1 represents a rating lower than the low vignette, the value 2 represents a rating at the level of the low vignette, the value 3 represents a rating higher than the low but lower than the medium vignette, and so forth. The maximum score, 7, is assigned when a student’s self-reported response is higher than the rating of the high vignette. In other words, low values are assigned when a self-report rating is relatively low compared to the evaluation of the vignettes, and high values are assigned when a self-report rating is relatively high compared to the evaluation of the vignettes. In this way, the three vignettes are used to anchor student judgements, providing context for the ratings on other questions sharing the same response scale. Scoring is applied on the individual student level using each student’s responses to the vignettes as an anchor for this student’s self-reported responses to various Likert-type questions. Table 17.1 illustrates the differences in possible values assigned to original and anchored item responses.

Table 17.1 Possible values for original and anchored item responses

Responses to question as presented in questionnaire	Strongly disagree	Disagree	Agree	Strongly agree			
	1	2	3	4			
Anchored responses	**Lower than low vignette**	**Same as low vignette**	**In between low and medium vignette**	**Same as medium vignette**	**In between medium and high vignette**	**Same as high vignette**	**Higher than high vignette**
	1	2	3	4	5	6	7

A graphical illustration of the scoring procedure based on vignettes for three examples with and without ties is given in Figure 17.1. The three hypothetical students in this example provided exactly the same responses to the three self-reported items shown, but differ in their responses to the vignettes. As a result, scores on the anchored items also differ between the three students.

Clear interpretation of the vignettes in terms of the relative ordering of low, medium, and high levels of the described characteristics was one requirement for the use of vignettes. Two special cases are given when there are ties in the responses to the anchoring vignettes (i.e. a student chooses the same agreement category for two or all three vignettes) or when responses to the anchoring vignettes violate the expected order of vignettes (i.e. a student chooses a higher agreement category for a vignette representing a low value on the underlying construct than for a vignette representing a high value on the underlying construct). The scoring method used in PISA 2012 addresses these two cases in the following way.

Tied anchor evaluations

The anchoring vignette items administered in the student questionnaire used a 4-point Likert scale. Students had to place three vignettes on this scale. Not all students chose distinct values for each vignette, but assigned the same value to two or sometime all three vignettes instead. Several different approaches how to treat these cases with ties in the vignette rankings were investigated (Bertling and Kyllonen, 2013). A “lower bound” scoring seemed most promising based on these investigations. During scoring, “lower bound scores” were assigned to students with tied vignette responses. For example, if a student evaluated the teachers with low and medium levels of classroom management identical, scores were adjusted based on the lower level. Choosing the lower bound score created the largest variation and allowed for the best differentiation between individuals. Also, results show that using lower bound scores maximises the criterion-correlations.

Order violations in anchor evaluations

If students do not rate vignettes in the expected order, there are several options as to how these cases can be treated. Based on initial Main Survey data four different approaches to treat order violations were tested to investigate whether including adjusted scores for these students might be possible in the international database or not. That is, these methods were compared against the most conservative treatment of order violations, namely exclusion of students with order violations prior to anchoring adjustments.

Methods A-C: This approach treats order violations by re-classifying "unallowed" vignette evaluations as ties. That is, if a student rates the highest vignette lower than the medium vignette, responses for this student would be rescaled in a way that the ratings for the medium and high vignette are tied. For instance, the rank order "low, high, med" would be rescaled into "low, {med, high}", with the brackets indicating that the same rank is assigned to the medium and high vignette. Table 17.2 presents all possible patterns of order violations and the respective rescaled response patterns. Note that, in most cases order violations are rescaled into complete ties of all vignettes (i.e., "{low, medium, high}"). Whenever ties are created in this post-hoc manner, there are at least two or three different variants how these ties can be created. For instance, in the example used above ("low, {med, high}") the tie could be created at the value the respondent assigned to the medium, or to the high vignette. All possible variants of creating ties were implemented based on initial Main Survey data for 52 countries. Method A created ties at the lowest possible value, method B created ties at the medium value, and method C created ties at the highest possible value. Details are provided in Table 17.2.

Method D: This approach treats order violations by essentially ignoring the predesignated vignette categories. That is, method D involves treating the vignette with the highest rating by an individual as the "high" exemplar, the one with the lowest rating as the "low" exemplar, and the one in the middle as the "middle" exemplar regardless of their predesignated vignette assignments. If the individual rated the lowest predesignated vignette as the highest, method D treats it as the highest. Thus order violations are transformed into non-violations. The rationale for this approach is as follows. The response style literature identifies a common phenomenon characterised as a tendency for respondents to respond to a rating scale in a way "irrespective of item content" (Fischer et al., 2009). Given this phenomenon, it should not be surprising if due to inattention, comprehension failures, or focusing on the wrong vignette description features, a respondent might mistakenly understand a predesignated "high" exemplar as a low one or a "low" exemplar as a high one. At the same time, the individual might be aware that exemplars vary in their standing on the dimension on which they are to be rated. Thus through the pattern of ratings respondents express their use of the rating scale (e.g., extreme, narrow, biased towards high, biased towards low), even when the respondent's ordering of ratings does not conform to the predesignated ordering of ratings. Following this logic, a respondent's self-rating can still be interpreted with respect to the vignette ratings, and mapped according to the non-metric remapping rules, even though the respondent did not interpret the vignettes in accordance with their predesignated categories.

Results for methods A-D were compared based on initial Main Survey data, especially their impact on the correlations with mathematics proficiency. Findings indicated considerable variation in achievement correlations depending on which method is used, with method C resulting in, on average, the largest correlations, followed by method D. Following discussions with the PISA Technical Advisory Group (TAG), method C was chosen as an appropriate re-scaling method for the international database. Under this method, if order violations in the vignette ratings are present, they are re-classified into ties. That is, if a student rates the highest vignette lower than the medium vignette, responses for this student would be re-scaled in a way that the ratings for the medium and high vignette are tied. For instance, the rank order "low, high, med" would be rescaled into "low, {med, high}", with the brackets indicating that the same rank is assigned to the medium and high vignette. Note that, in most cases, order violations are rescaled into complete ties of all vignettes (i.e. "{low, medium, high}"). Ties are created at the highest response category chosen by the student. For instance, in the example used above ("low, {med, high}") the tie is created at the value the respondent assigned to the high vignette. Ties are then analysed as described above. Table 17.2 shows the rescaling procedure for order violations for all four possible methods.

Comparison of original and anchored indices

Field Trial analyses and initial Main Survey analyses consistently showed that within-country correlations with achievement tended to be higher for anchored scales, and correlations at the between-student within-country level and the between-country level did not show the inconsistencies found for unanchored scales. No "paradoxical" correlations were found for any anchored index, but correlations on the between-country level were similar to student-level correlations, both within countries and for the pooled sample. The absolute values of the between-country correlations tended to be larger than the correlations at the between-student within-country level.

Table 17.2 Detailed rescaling specifications for the different methods to address order violations

				Recoding into ties									Re-ordering		
				Method A			Method B			Method C			Method D		
				Recoding of actual values, tied at lower			Recoding of actual values, tied at medium			Recoding of actual values, tied at high			Re-ordering or original responses		
Original response	Valid %	Type	Recoding into allowed type with ties	Low vignette	Medium vignette	High vignette	Low vignette	Medium vignette	High vignette	Low vignette	Medium vignette	High vignette	Low vignette	Medium vignette	High vignette
^421	0.04%	3,2,1	(1,2,3)	1	1	1	2	2	2	4	4	4	1	2	4
^422	0.05%	3,(1,2)	(1,2,3)	2	2	2	4	4	4	4	4	4	2	4	4
^342	0.05%	3,1,2	(1,2,3)	2	2	2	3	3	3	4	4	4	2	3	4
^342	0.05%	3,1,2	(1,2,3)	2	2	2	3	3	3	4	4	4	2	3	4
^321	0.05%	3,2,1	(1,2,3)	1	1	1	2	2	2	3	3	3	1	2	3
^242	0.07%	(1,3),2	(1,2,3)	2	2	2	2	2	2	4	4	4	2	2	4
^221	0.07%	3,(1,2)	(1,2,3)	1	1	1	2	2	2	2	2	2	1	2	2
^414	0.08%	2,(1,3)	(1,2,3)	1	1	1	4	4	4	4	4	4	1	4	4
^332	0.08%	3,(1,2)	(1,2,3)	2	2	2	3	3	3	3	3	3	2	3	3
^314	0.08%	2,1,3	(1,2,3)	1	1	1	3	3	3	4	4	4	1	3	4
^443	0.09%	3,(1,2)	(1,2,3)	3	3	3	4	4	4	4	4	4	3	4	4
^432	0.09%	3,2,1	(1,2,3)	2	2	2	3	3	3	4	4	4	2	3	4
^212	0.12%	2,(1,3)	(1,2,3)	1	1	1	2	2	2	2	2	2	1	2	2
^413	0.12%	2,3,1	(1,2,3)	1	1	1	3	3	3	4	4	4	1	3	4
^232	0.13%	(1,3),2	(1,2,3)	2	2	2	2	2	2	3	3	3	2	2	3
^431	0.13%	3,2,1	(1,2,3)	1	1	1	3	3	3	4	4	4	1	3	4
^213	0.14%	2,1,3	(1,2,3)	1	1	1	2	2	2	3	3	3	1	2	3
^132	0.14%	1,3,2	1,(2,3)	1	2	2	1	3	3	1	3	3	1	2	3
^214	0.15%	2,1,3	(1,2,3)	1	1	1	2	2	2	4	4	4	1	2	4
^343	0.15%	(1,3),2	(1,2,3)	3	3	3	3	3	3	4	4	4	3	3	4
^142	0.15%	1,3,2	1,(2,3)	1	2	2	1	4	4	1	4	4	1	2	4
^322	0.18%	(2,3),1	(1,2,3)	2	2	2	2	2	2	3	3	3	2	2	3
^422	0.20%	(2,3),1	(1,2,3)	2	2	2	2	2	2	4	4	4	2	2	4
^313	0.21%	2,(1,3)	(1,2,3)	1	1	1	3	3	3	3	3	3	1	3	3
^331	0.23%	3,(1,2)	(1,2,3)	1	1	1	3	3	3	3	3	3	1	3	3
^243	0.25%	1,3,2	1,(2,3)	2	3	3	2	4	4	2	4	4	2	3	4
^312	0.26%	2,3,1	(1,2,3)	1	1	1	2	2	2	3	3	3	1	2	3
^143	0.41%	1,3,2	1,(2,3)	1	3	3	1	4	4	1	4	4	1	3	4
^412	0.42%	2,3,1	(1,2,3)	1	1	1	2	2	2	4	4	4	1	2	4
^433	0.42%	(2,3),1	(1,2,3)	3	3	3	3	3	3	4	4	4	3	3	4
^424	0.47%	2,(1,3)	(1,2,3)	2	2	2	4	4	4	4	4	4	2	4	4
^441	0.50%	3,(1,2)	(1,2,3)	1	1	1	4	4	4	4	4	4	1	4	4
^423	0.52%	2,3,1	(1,2,3)	2	2	2	3	3	3	4	4	4	2	3	4
^411	0.54%	(2,3),1	(1,2,3)	1	1	1	1	1	1	4	4	4	1	1	4
^121	0.54%	(1,3),2	(1,2,3)	1	1	1	1	1	1	2	2	2	1	1	2
^434	0.57%	2,(1,3)	(1,2,3)	3	3	3	4	4	4	4	4	4	3	4	4
^231	0.60%	3,1,2	(1,2,3)	1	1	1	2	2	2	3	3	3	1	2	3
^341	0.79%	3,1,2	(1,2,3)	1	1	1	3	3	3	4	4	4	1	3	4
^323	0.93%	2,(1,3)	(1,2,3)	2	2	2	3	3	3	3	3	3	2	3	3
^324	1.05%	2,1,3	(1,2,3)	2	2	2	3	3	3	4	4	4	2	3	4
^311	1.55%	(2,3),1	(1,2,3)	1	1	1	1	1	1	3	3	3	1	1	3
^211	1.72%	(2,3),1	(1,2,3)	1	1	1	1	1	1	2	2	2	1	1	2
^131	3.03%	(1,3),2	(1,2,3)	1	1	1	1	1	1	3	3	3	1	1	3
^241	4.55%	3,1,2	(1,2,3)	1	1	1	2	2	2	4	4	4	1	2	4
^141	5.64%	(1,3),2	(1,2,3)	1	1	1	1	1	1	4	4	4	1	1	4

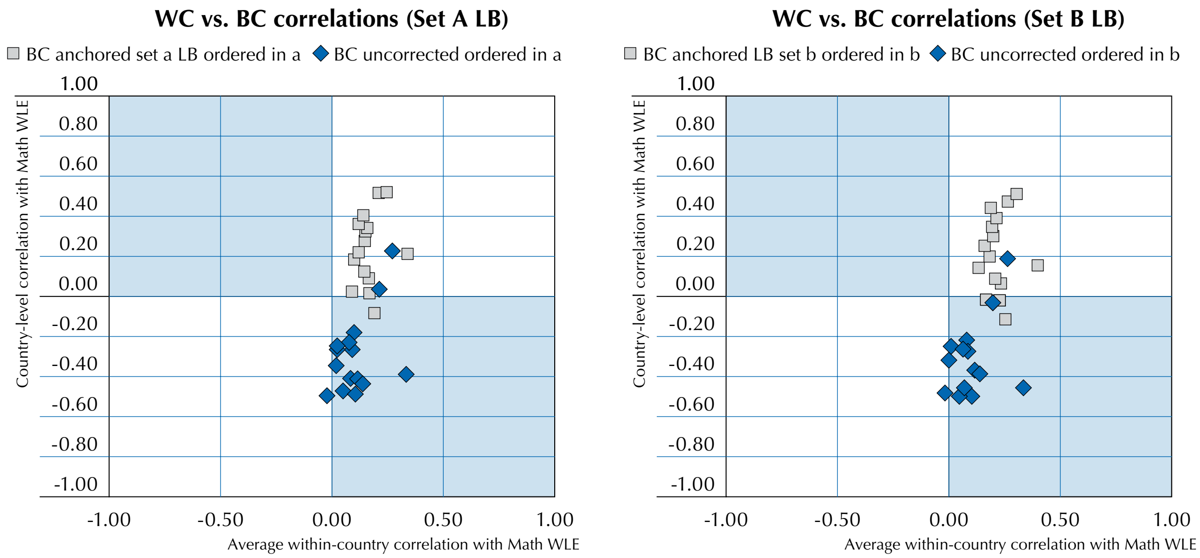
Note: Relative frequencies here are based on initial Main Survey data from 52 countries. Shaded fields indicate responses patterns that only partly violate the correct order of vignettes.

Comparisons of the two sets of anchoring vignettes showed very similar results with no major differences in the pattern of correlations between scales and achievement (see Figure 17.2 for a graphical illustration). Further, in comparison with the original indices, anchored indices showed smaller degrees of Differential Item Functioning (DIF) across countries and smaller correlations with indicators of acquiescence or disacquiescence response styles (Bertling and Kyllonen, 2013).

These findings supported the validity of the anchoring corrections and suggested that meaningful relationships of background indexes with achievement were suppressed by the dominance of general response tendencies among students when uncorrected responses were scaled.

■ Figure 17.2 ■

Alignment of average within-country and between-country correlations for unadjusted and anchored indexes



Note: In this figure, "LB" stands for lower bound and "Math WLE" stands for Mathematic: Weighted Least Square Estimate.

Four possible scenarios for making anchoring adjustments in the international database were presented to the PISA Technical Advisory Group (Bertling and Kyllonen, 2013). In scenario 1, only the two indices, namely Teacher Support (MTSUP) and Classroom Management (CLSMAN) would be anchored, as these scales directly map to the content of the vignettes. This scenario represented a construct specific anchoring approach. Scenarios 2-4 represent a construct non-specific approach. All questionnaire items with the same response format ("strongly agree" to "strongly disagree") can be anchored under these scenarios. Scenarios 2-4 differed in terms of which set of vignettes is used for the adjustment, or whether both sets are used. Results for the different scenarios were discussed with the PISA Technical Advisory Group (TAG). Anchoring adjustments for 12 Student Questionnaire scales were thereafter incorporated into the international database (see Chapter 16 for further details of all Student Questionnaire indices).

Limitations and cautions

The alternative scoring approach for Likert-type items based on vignettes makes the frame of reference for scoring of questionnaires items more transparent and can thereby help in interpreting students' responses across different countries and education systems. There are, however, several assumptions that underlie the use of anchoring vignettes in the context of PISA. Caution is advised when interpreting adjusted indices using anchoring vignettes (Bertling and Kyllonen, 2013).

The scoring approach is based on two main identifying assumptions: "vignette equivalence" and "response consistency" (e.g., Van Soest et al., 2011). First, the vignette equivalence assumption posits that different respondents interpret the vignette scenario in the same way. In other words, there is an assumption that all differences in the ratings of the vignettes should be attributable to the differences in how respondents interpret and use the agreement scale, but not to the differences in how respondents interpret the vignette scenario themselves. Second, the response consistency assumption posits that respondents use the same standards both in evaluating themselves and in providing an evaluation of the vignette scenario.

In this context, it should be noted that the original anchoring vignette method was developed to anchor items relating to the content described in the scenario of the anchoring vignette only. In the PISA 2012 Student Questionnaire, this method was extended so that the same vignette scenario was applied to a large set of different items. This extension was possible because of an assumption that an individual's rating standards are invariant across different contexts whenever the same rating scale is used. This means that students are expected to use a four-point Likert scale with the categories "Strongly disagree" to "Strongly agree" in a reasonably comparable way for the different questions included in the Student Questionnaire, whether these refer to items such as "I learn mathematics quickly" or items such as "My teacher helps students with their learning".

The scoring process anchoring student responses using vignette scenarios depends on the particular vignette scenarios (i.e. where on the continuum of the underlying construct the vignettes are located) and the number of vignettes used. While PISA 2012 data suggests reasonable consistency of results across the two sets of vignettes, further research is needed to fully understand the effects of different vignette contexts and how the validity of results depends on the number of vignettes and number of scale points used. For instance, gains in validity might be larger for questions that capture similar constructs as the constructs described in the vignettes.

The order of vignettes and self-reports in the questionnaire may have an influence on the results. As Hopkins and King (2010) showed, administering vignettes first might have a priming effect that reduces inter-individual differences in interpretation of the response scale. In the PISA 2012 Student Questionnaire some self-reported questions using the four-point Likert scale are presented before the vignettes and others are asked after the vignettes.

Finally, in order to use data from all students, including students with tied anchor evaluations (e.g. students who give the same ratings for two vignettes classified as low and medium) or "order violations" (e.g. students who give lower ratings to a vignette classified as high as to a vignette classified as medium or low), additional assumptions are needed, as described in the previous sections. Future research is needed to fully understand students' response processes.

It is recommended that adjusted indices using anchoring vignettes should be interpreted in addition to classical indices, not as a replacement. Both values on classical questionnaire indices and on adjusted indices can be influenced by students' systematic or unsystematic response behaviours. Examining both of these indices provides a basis for a more general picture of relationships and effects that is less tied to a single survey method only.

Topic familiarity with signal detection correction

Topic familiarity with signal detection correction of "Overclaiming Technique" (OCT; Paulhus, Harms, Bruce and Lysy, 2003; see also Zimmerman, Broder, Shaughnessy and Underwood, 1977), was used in the PISA 2012 as one way to enable adjustments for differences in response tendencies (see Chapter 3).

In the PISA 2012 Student Questionnaire an OCT was operationalised by asking students to indicate their familiarity – on a 5-point scale from "never heard of it" to "know it well; understand the concept" – with actual mathematics concepts (e.g. "polynomial function") and foils (e.g. "proper number"). Foils were created by combining a term from grammar (i.e. proper, as in proper noun; subjunctive, as in subjunctive mood; declarative as in declarative sentence) with a mathematical term (i.e. number; scaling; fraction, respectively). The questionnaire items are displayed in Figure 17.3.

Two indexes were computed from students' responses. One was a simple mean of their familiarity scores on the 5-point scale (FAMCON), and the other took that mean and subtracted from it the mean familiarity score of the foil concepts (FAMCONC).

Field Trial and initial Main Survey analyses (Bertling and Roberts, 2011) showed increased cross-cultural comparability of correlations with achievement for the adjusted compared to the unadjusted index. Also, familiarity ratings for the foil concepts only were strongly negatively correlated with achievement. Based on consultation with the PISA 2012 Technical Advisory Group who discussed preliminary findings from both Field Trial and Main Survey data at two meetings in 2012 and 2013, the adjusted topic familiarity index (FAMCONC) was included in the international database in addition to the original unadjusted index. While the overclaiming correction was applied to one index only in the international database, this correction might be applicable to other indices as well to correct for general response styles (Kyllonen, Bertling and Roberts, 2013).

Situational Judgement Tests

As discussed in Chapter 3, Situational Judgement Test items (SJTs; see Weekley and Ployhart, 2006, for an overview) were included in the PISA 2012 questionnaire as another possible way to adjust for differences in response behaviour.

The problem solving SJT in the PISA 2012 Student Questionnaire consisted of three different scenarios that described situations that could arise in the course of solving a problem. Questions focus on a person's initial response to a problem as well as possible approaches to take if one's initial response to the problem fails. The three scenarios involved a) a problem with a text message on a mobile phone, b) route selection for getting to a zoo and c) a malfunctioning ticket vending machine. Response options to each scenario tapped into different problem-solving strategies, namely systematic strategies, unsystematic strategies and seeking help.

■ Figure 17.3 ■

Topic familiarity items in the PISA 2012 Student Questionnaire (ST62)

ST62

Q **Thinking about mathematical concepts: how familiar are you with the following terms?**

(Please tick only one box in each row.)

	Never heard of it	*Heard of it once or twice*	*Heard of it a few times*	*Heard of it often*	*Know it well, understand the concept*
a) Exponential Function	☐1	☐2	☐3	☐4	☐5
b) Divisor	☐1	☐2	☐3	☐4	☐5
c) Quadratic Function	☐1	☐2	☐3	☐4	☐5
d) <Proper Number>	☐1	☐2	☐3	☐4	☐5
e) Linear Equation	☐1	☐2	☐3	☐4	☐5
f) Vectors	☐1	☐2	☐3	☐4	☐5
g) Complex Number	☐1	☐2	☐3	☐4	☐5
h) Rational Number	☐1	☐2	☐3	☐4	☐5
i) Radicals	☐1	☐2	☐3	☐4	☐5
j) <Subjunctive Scaling>	☐1	☐2	☐3	☐4	☐5
k) Polygon	☐1	☐2	☐3	☐4	☐5
l) <Declarative Fraction>	☐1	☐2	☐3	☐4	☐5
m) Congruent Figure	☐1	☐2	☐3	☐4	☐5
n) Cosine	☐1	☐2	☐3	☐4	☐5
o) Arithmetic Mean	☐1	☐2	☐3	☐4	☐5
p) Probability	☐1	☐2	☐3	☐4	☐5

Scenarios and items were designed along the goals of PISA that the PISA 2012 problem solving assessment should be based on problem solving in authentic concrete contexts rather than with abstract tasks (OECD, 2013a). A four point response format with the options “I would definitely not do this”, “I would probably not do this”, “I would probably do this”, and “I would definitely do this” was chosen. This is consistent with other background items in the PISA Student Questionnaires, where a 4-point response scale is the most frequently applied scale. Figure 17.4 shows an example item.

Based on exploratory factor analyses as well as more complex multi-trait multi-method models, three dimensions were identified in the SJT data associated with (a) systematic problem-solving behaviours (i.e., behaviours resulting from an analysis of the problem situation and planful acting), (b) unsystematic problem solving behaviours (i.e., unplanful and not the result of a thorough analysis of the problem situation but rather impulsive behavioural tendencies that are unrelated to the specific needs of the situation), and (c) help seeking behaviours (i.e., tendencies to rely on others’ knowledge and expertise regarding how to solve the problem; see Kyllonen, Bertling and Roberts 2013,

for details). While the three-factor structure was found across most countries, reliabilities of the possible indices were low (Cronbach's $\alpha < .70$). This is a typical finding for situational judgement tests that has been described in the literature many times (e.g., Christian et al., 2010). It was, however, decided not to include these indices in the international database as the internal consistencies fell short of the standards for internal consistencies of indices used in PISA.

■ Figure 17.4 ■

Example situational judgement test item from the PISA 2012 student questionnaire (ST96)

ST96

Q *Suppose that you have been sending text messages from your mobile phone for several weeks. Today, however, you can't send text messages. You want to try to solve the problem.*

What would you do? For each suggestion, tick the option that best applies to you.

(Please tick only one box in each row.)

	I would definitely do this	*I would probably do this*	*I would probably not do this*	*I would definitely not do this*
a) I press every button possible to find out what is wrong.	❑ 1	❑ 2	❑ 3	❑ 4
b) I think about what might have caused the problem and what I can do to solve it.	❑ 1	❑ 2	❑ 3	❑ 4
c) I read the manual.	❑ 1	❑ 2	❑ 3	❑ 4
d) I ask a friend for help.	❑ 1	❑ 2	❑ 3	❑ 4

Forced Choice

The Forced-Choice (FC) item format was another way in which PISA 2012 sought to overcome measurement issues related to the use of traditional Likert-type response formats (see Chapter 3).

Five items used the forced-choice format to measure students' plans regarding mathematics at some stage in the future (*MATINTFC*) in PISA 2012. The first item had students decide between taking additional courses in mathematics or the language of the test after school finished. The second item asked whether students planned on majoring in a subject requiring mathematics or science at college (or equivalent educational institution in different countries). The third item asked whether students were willing to study harder in either their mathematics classes or in classes teaching the language of the test. For the fourth item, respondents had to indicate whether they were planning on taking as many mathematics or science classes as possible during their education. The fifth and last item of that battery required respondents to choose whether they were planning on pursuing a career that involved a lot of mathematics or science.

All items were reversed so that respondents who chose mathematics over either science or the test language were allocated a higher code. Then, responses to all five items were IRT (Item Response Theory) scaled and written onto the international database as a variable indicating mathematics intentions (*MATINTFC*).

One index based on the forced choice items is included in the international database.

■ Figure 17.5 ■

Example forced choice item from PISA 2012 Student Questionnaire (ST48e)

ST48

Q **For each pair of statements, please choose the item that best describes you.**

e) *Please tick only one of the following two boxes.*

❑1 I am planning on pursuing a career that involves a lot of mathematics.

❑2 I am planning on pursuing a career that involves a lot of science.

LESSONS CONCERNING ONLINE DELIVERY OF SCHOOL QUESTIONNAIRES

In addition to paper-based delivery of the questionnaires, PISA 2012 introduced online administration for the school questionnaire. On this first occasion online administration of the school questionnaire was optional.

Nineteen PISA participants took up the Online School Questionnaire option in the Main Survey in PISA 2012 which resulted in the administration of the questionnaire in 24 language versions. Participants included: Australia, Austria, Chile, Cyprus,[1] Denmark, Estonia, Finland, Hungary, Iceland, Ireland, Israel, Korea, Liechtenstein, Norway, Portugal, Singapore, Slovenia, Switzerland and Chinese Taipei.

Based on feedback from National Centres after the PISA 2012 Main Survey, the international contractor suggests that future cycles of PISA should consider how to further integrate the simultaneous processing of both paper-based and online questionnaires and materials during questionnaire development, negotiation, online authoring, verification, implementation and data management and cleaning.

While most processes were more efficient when compared to the paper-based administration, some issues specific to the Online School Questionnaire needed to be considered when interpreting results – particularly with regard to percentages of missing values for Online School Questionnaire data.

Questions SC14, SC16, SC18, SC22, SC25, SC33, SC34, and SC39 in the School Questionnaire contained more than six items. This required respondents to use a scroll-bar to view all items as the response scale for these items was fixed and had to always be visible. As such, higher levels of missing data for later items in 'long' questions were probable.

According to an analysis of Main Survey Online School Questionnaire data, on average, three to five per cent of respondents consistently failed to scroll-down for all later items, namely any item after the first six items, in these questions. Other missing values for items in these questions were randomly distributed, as respondents used the scroll-bar for some questions but failed to do so for others.

PISA participants that had more than five schools with missing values for all later items for SC14, SC16, SC18, SC22, SC25, SC33, SC34, and SC39 were given the opportunity to re-open their Online School Questionnaire to collect the missing data. Of the six participants that were approached two decided to re-open their questionnaires.

To address the question as to how the online administration mode compares to the paper-based administration mode, the average proportion of missing (code "99") across all questions in the School Questionnaire was computed for each country. Results are shown in Table 17.3.

Table 17.3 Average proportion of missing data across all questions in the School Questionnaire

CNT	online	% missing	CNT	online	% missing	CNT	online	% missing
ALB	no	2.91	GRC	no	3.18	NOR	yes	3.71
ARE	no	8.31	HKG	no	3.03	NZL	no	2.70
ARG	no	6.60	HRV	no	0.98	PER	no	4.60
AUS	yes	3.14	HUN	yes	1.62	POL	no	2.08
AUT	yes	5.76	IDN	no	3.99	PRT	yes	2.37
BEL	no	4.55	IRL	yes	4.43	QAT	no	5.02
BGR	no	4.03	ISL	yes	3.89	QCN	no	0.85
BRA	no	8.38	ISR	yes	3.53	ROU	no	0.22
CAN	no	3.55	ITA	no	4.67	RUS	no	1.57
CHE	yes	2.85	JOR	no	5.63	SGP	yes	1.84
CHL	yes	1.58	JPN	no	0.17	SRB	no	4.55
COL	no	7.02	KAZ	no	1.57	SVK	no	1.76
CRI	no	4.01	KOR	yes	1.09	SVN	yes	1.96
CYP*	yes	3.09	LIE	yes	3.17	SWE	no	2.79
CZE	no	3.60	LTU	no	2.46	TAP	yes	2.00
DEU	no	3.31	LUX	no	4.32	THA	no	0.81
DNK	yes	3.80	LVA	no	3.61	TUN	no	5.38
ESP	no	4.69	MAC	no	0.16	TUR	no	4.39
EST	yes	2.25	MEX	no	4.51	URY	no	1.20
FIN	yes	2.50	MNE	no	1.39	USA	no	4.91
FRA	no	4.36	MYS	no	4.39	VNM	no	0.71
GBR	no	2.32	NLD	no	2.77			

* See note 1 at the end of this chapter.

As can be seen in Table 17.3, the average proportion of missing data across all questions in the online mode ranged from 1.09% in Korea to 5.76% in Austria. For the paper-based School Questionnaire, average proportions of missing data ranged from 0.16% in Macao-China to 8.38% in Brazil. A comparison of the average proportion of missing data for online mode and paper-based mode of administrations revealed no substantive differences between the two modes.

Another way of exploring possible differences in missing data between administration modes is to look at individual items in the School Questionnaire. To this end, the items with the lowest and highest average proportion of missing data across countries were examined. Across both online and paper-based countries, the item with the lowest average proportion of missing data (0.17%) was recorded for the question that asked principals to indicate the size of the community in which their school was located (SC03Q01). For the online countries, the average proportion of missing data was 0.16% across the three online countries that had missing data for this question. For the paper-based countries, the average proportion of missing was very similar at 0.17% with ten of the 47 countries recording missing data. One of the highest average proportions of missing data across countries (18.9%) was recorded for SC10Q22 which asked about the number of part-time mathematics teachers with <ISCED 5A> qualification. The average proportion of missing data across the 19 countries which administered this question online was 5.62% while the average across the 46 countries which administered this version in paper-based mode was 24.89%. This difference indicated a considerably lower proportion of missing data for this item in the online mode compared with the paper-based mode (see Tables 17.13, 17.14 and 17.15 available at *www.oecd.org/pisa*).

Improvements in online administration of surveys for future cycles of PISA could consider the network response times for respondents that are geographically farther away from the server. With the server being located in Europe, participants outside this continent reported more problems with log-in and loading times for the questionnaire and between screens.

Lastly, future online questionnaire administration could consider how to improve questionnaire administration for respondents with widely differing local operating environments. Often, respondents were unable to update local browsers and operating systems due to centralised restrictions placed upon school electronic information systems. Future cycles of questionnaire administration should further consider how questionnaire and platform development as well as data management can accommodate the local operating environments of respondents.

THE EFFECT OF ISCO-08 COMPARED WITH ISCO-88 ON ESCS AND PERFORMANCE

In the PISA 2012 Main Survey, all questions relevant to the computation of ESCS (the *PISA index of economic, social and cultural status*) were included in the common part of the Student Questionnaire, which was administered to all students (see also the section on the rotated Student Questionnaire design Chapter 3).

In a first step, the International Standard Classification of Occupations in their 2008 version (ISCO-08) codes were mapped back onto the ISCO codes in their 1988 (ISCO-88) version. In addition, the most recent transformation from ISCO-08 to ISEI-08 published on Harry Ganzeboom's website *http://home.fsw.vu.nl/hbg.ganzeboom/isco08/* [2] was used.

ESCS was calculated using the resultant International Socio-Economic Index of occupational status (ISEI-08) codes in the same way as detailed in the *PISA 2006 Technical Report* (OECD, 2009), with straightforward changes due to differently named variables on the scale measuring home possessions (HOMEPOS). The computation for this comparison is slightly different from that used for the official database, so the results are not identical. In Table 17.4, the mean ESCS values using ISCO-88 and ISCO-08 are given for 64 of the 65 countries that participated in PISA 2012 as problems relating to the questionnaire data required for ESCS occurred in Albania. The first column shows the country. The next four columns show the means and standard deviations (S.D.) for the two ESCS versions using ISCO-88 (columns 2-3) and using ISCO-08 (columns 4-5). The next two columns report the absolute difference in the means between the two ESCS versions (column "Abs Diff") and the correlation between the two ESCS versions (column "Corr.").

It can be seen that the magnitude of the absolute differences is small with an average across all countries of 0.02, with 47 of the 64 countries in the analysis having a difference lower or equal to this value. The two largest absolute differences in the means are recorded for Jordan (0.07) and United Arab Emirates (0.06). The lowest correlations can be observed for Japan, Qatar and Spain with values of 0.94, 0.93 and 0.93 respectively. All other correlations are above 0.95.

As another way of examining the implication of the change in ISCO coding, effects of ESCS on student performance were investigated. In PISA 2012, the effect of ESCS on mathematics performance in a regression analysis ranged from 17 PISA points for Macao-China to 58 PISA points for Chinese Taipei (OECD, 2013b). This corresponds to a change in achievement of this number of PISA points for each unit change in ESCS. So, taking the first coefficient of 17 as an example, if a country has an average mathematics performance of 500 PISA points, one unit increase (=+1.0) in ESCS would correspond to an average performance of 517.

The average absolute change in mean ESCS for the countries is 0.02, which, depending on the size of the regression coefficient – from about 17 to 58 in mathematics achievement in PISA 2012 – would lead to a shift in the value of ESCS associated with a certain level of achievement. As the standard deviations are almost unchanged, only the intercept of the regression would change. For the two countries with the greatest change in ESCS, this would have the following implications. For Jordan, in mathematics in 2012, with a score-point difference associated with one unit on the ESCS of 22 the 0.07 change leads to about 22*0.07=1.6 PISA points change in the intercept, whereas the United Arab Emirates with a regression coefficient of 33 together with a change in ESCS of 0.06 would lead to a change of 33*0.06=2.0 PISA points.

Hence, the country ESCS means would seem not to have been greatly affected by the change in the ISCO and subsequent ISEI coding.

To understand better how the distribution of ESCS was affected, percentile plots of the two ESCS versions were performed. The circles in the plots below show the plot of the 1st, 2nd to 99th percentiles points of the two ESCS distributions. The fitted spline line is a piecewise defined low-order polynomial.

Figure 17.6 shows a percentile plot of the ESCS using the new version of ISEI-08 ("escsP") against the ESCS using ISEI-88 ("escs_88P") for Hong-Kong China as a case with a relatively higher correlation between the two measures using PISA 2012 data from the Main Survey. The results show very little nonlinearity.

Figure 17.7 shows the same plot for Qatar, which has a slightly lower, but still high, correlation between the two ESCS versions than other countries. A small amount of curvature can be observed at the top end but no serious indication of nonlinearity.

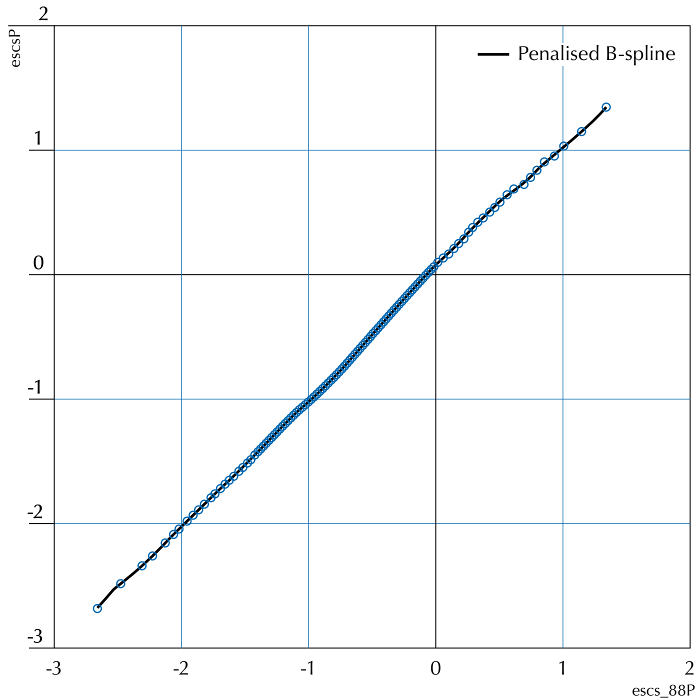
Table 17.4 ESCS calculated with alternative ISCO codings for PISA 2012 Main Survey

	ESCS using ISCO-88		ESCS using ISCO-08 (mapping)		Difference between the two ESCS versions	
CNT	Mean	SD	Mean	SD	Abs Diff	Corr.
ARE	0.40	0.87	0.45	0.84	0.06	0.95
ARG	-0.64	1.10	-0.62	1.10	0.02	0.97
AUS	0.23	0.80	0.23	0.79	0.00	0.95
AUT	0.12	0.85	0.10	0.86	0.02	0.97
BEL	0.15	0.91	0.15	0.91	0.00	0.97
BGR	-0.17	1.09	-0.15	1.04	0.02	0.97
BRA	-1.12	1.21	-1.13	1.19	0.02	0.98
CAN	0.41	0.86	0.41	0.86	0.00	0.97
CHE	0.19	0.90	0.17	0.89	0.02	0.97
CHL	-0.58	1.14	-0.58	1.14	0.00	0.98
COL	-1.16	1.18	-1.17	1.19	0.01	0.98
CRI	-0.90	1.22	-0.88	1.24	0.03	0.98
CYP*	0.21	0.89	0.19	0.89	0.01	0.97
CZE	-0.04	0.73	-0.05	0.73	0.01	0.96
DEU	0.24	0.93	0.21	0.93	0.02	0.97
DNK	0.39	0.82	0.43	0.83	0.03	0.96
ESP	-0.19	1.03	-0.19	1.03	0.00	0.93
EST	0.14	0.80	0.14	0.79	0.01	0.97
FIN	0.32	0.77	0.35	0.77	0.03	0.96
FRA	-0.04	0.80	-0.04	0.79	0.00	0.96
GBR	0.24	0.80	0.28	0.80	0.04	0.95
GRC	-0.05	0.98	-0.06	1.00	0.01	0.98
HKG	-0.68	0.93	-0.67	0.96	0.01	0.97
HRV	-0.18	0.84	-0.22	0.84	0.04	0.96
HUN	-0.22	0.94	-0.24	0.95	0.03	0.97
IDN	-1.73	1.07	-1.76	1.11	0.03	0.98
IRL	0.10	0.84	0.11	0.84	0.01	0.97
ISL	0.75	0.82	0.78	0.82	0.03	0.95
ISR	0.13	0.88	0.16	0.87	0.02	0.96
ITA	0.00	0.98	-0.02	0.96	0.02	0.97
JOR	-0.37	1.03	-0.30	1.01	0.07	0.96
JPN	-0.09	0.69	-0.12	0.69	0.03	0.94
KAZ	-0.22	0.78	-0.21	0.77	0.01	0.96
KOR	-0.08	0.72	-0.03	0.73	0.05	0.95
LIE	0.44	0.92	0.40	0.90	0.04	0.97
LTU	-0.01	0.93	0.00	0.91	0.01	0.97
LUX	0.10	1.12	0.11	1.09	0.01	0.98
LVA	-0.13	0.93	-0.14	0.89	0.01	0.97
MAC	-0.77	0.85	-0.77	0.86	0.00	0.97
MEX	-1.16	1.31	-1.16	1.29	0.00	0.99
MNE	-0.12	0.89	-0.14	0.90	0.01	0.95
MYS	-0.67	0.96	-0.63	1.00	0.05	0.97
NLD	0.24	0.78	0.23	0.77	0.01	0.96
NOR	0.44	0.79	0.42	0.76	0.02	0.96
NZL	0.01	0.83	0.03	0.82	0.02	0.95
PER	-1.13	1.25	-1.15	1.24	0.02	0.99
POL	-0.17	0.86	-0.18	0.88	0.01	0.97
PRT	-0.45	1.20	-0.45	1.19	0.00	0.98
QAT	0.50	0.90	0.52	0.91	0.02	0.93
QCN	-0.23	0.94	-0.23	0.95	0.00	0.98
ROU	-0.36	0.96	-0.36	0.95	0.01	0.97
RUS	-0.01	0.78	0.01	0.76	0.02	0.95
SGP	-0.16	0.91	-0.14	0.92	0.02	0.97
SRB	-0.20	0.93	-0.19	0.90	0.01	0.98
SVK	-0.12	0.92	-0.16	0.92	0.04	0.97
SVN	0.11	0.87	0.10	0.87	0.02	0.97
SWE	0.25	0.82	0.26	0.81	0.01	0.96
TAP	-0.28	0.83	-0.30	0.83	0.02	0.96
THA	-1.21	1.17	-1.25	1.19	0.03	0.98
TUN	-1.15	1.28	-1.12	1.31	0.02	0.98
TUR	-1.47	1.10	-1.48	1.12	0.01	0.98
URY	-0.77	1.14	-0.78	1.14	0.01	0.98
USA	0.15	0.98	0.16	0.98	0.01	0.96
VNM	-1.71	1.13	-1.76	1.18	0.05	0.99
Average	**-0.23**	**0.94**	**-0.22**	**0.94**	**0.02**	**0.97**

*See note 1 at the end of this chapter.

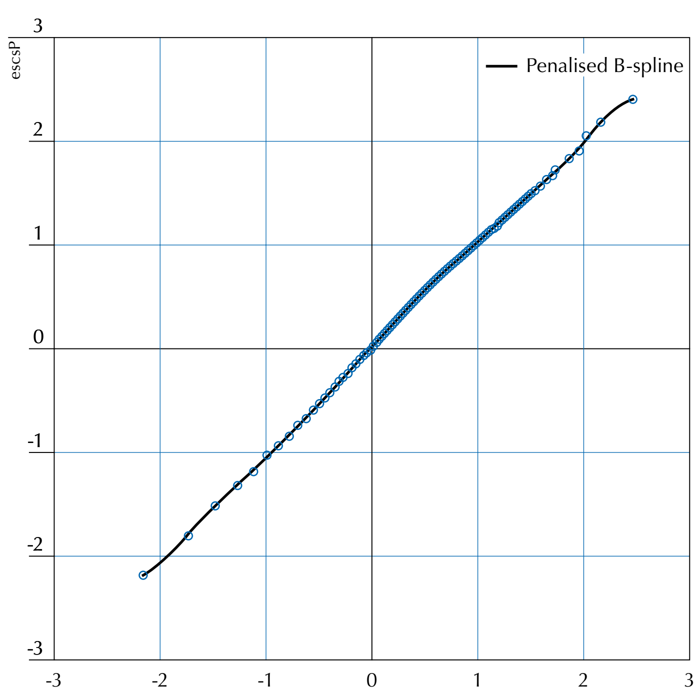
■ Figure 17.6 ■

Percentile plot of two ESCS versions – Hong Kong-China PISA 2012 Main Survey data



■ Figure 17.7 ■

Percentile plot of two ESCS versions – Qatar PISA 2012 Main Survey data



Percentile plots of the two ESCS versions were generated for all countries. Similar to Figure 17.7, plots seemed to show some flattening at the top end of the distributions for some countries, but, in general, the plots had little curvature as would be expected with the high correlations observed.

As the standard deviations for ESCS have not changed greatly in most countries, the effect of the update from ISCO-88 to ISCO-08 will be a shift of the intercept of the regression by up to 2 points of ESCS on achievement without changing the regression coefficient in most instances. Still, the analyses show a change in standard deviation of ESCS in a few countries. For example the standard deviation of ESCS in Indonesia increases from 1.07 to 1.11. This translates to a decrease in the coefficient by a factor of 1.07/1.11=0.96. In other words, instead of the regression coefficient taking on a value of 20 for ESCS in mathematics performance for Indonesia using the ISCO-08 coding the value would have been 21 using the ISCO-88 coding, which is not a large difference.

In the international database for PISA 2012, ESCS has been equated back to the earlier cycles using the 2012 ESCS as the base to establish trends rather than attempting to equate back to PISA 2009. This required the recomputing of ESCS using new ISCO-08 to ISEI-08 mapping for all cycles. Further details regarding the computation of ESCS are provided in Chapter 16.

An examination of the application of the new ISCO-08 code in the field

An analysis of the double coded mother's and father's occupation data from the Student Questionnaire administered as part of the PISA 2012 Field Trial (FT12) was conducted to investigate whether or not the change to the classification has impacted on the degree of agreement between the two coders.

In the FT12, two independent coders assigned a four digit ISCO code to mother's and father's occupation for a randomly selected number of students in 18 countries. A data file was produced with this information totalling 8,519 cases. Table 17.5 below shows the number of cases that were double coded for each country.

Table 17.5 Cases in the analysis of the PISA 2012 Field Trial double coding

CNT	Frequency	Percentage
BRA	321	3.8
CHE	94	1.1
CHL	1086	12.7
COL	152	1.8
DNK	224	2.6
HRV	1092	12.8
ISL	61	0.7
ISR	397	4.7
MAC	2609	30.6
MEX	300	3.5
PRT	451	5.3
RUS	227	2.7
SGP	222	2.6
SVN	292	3.4
SWE	187	2.2
TAP	263	3.1
TUR	342	4.0
URY	199	2.3
Total	**8519**	**100.0**

Cases were selected where the ISCO codes for the four-digit comparison differed between the two coders whereby missing responses for one or both coders were excluded. Differences were observed for (355) 4.2% of cases for mother's occupation and (427) 5.0% of cases for father's occupation. These data were tabulated to review which ISCO codes might be causing discrepancies between coders.

As getting the first two of the four digit codes correct achieves 90% accuracy (see Ganzeboom, 2010), cases that matched on the first two codes were considered to coincide sufficiently and removed. This left 2.4% of mother coded occupations and 3.3% of father coded occupations with discrepancies that were further examined.

From an examination of this small number of cases, no obvious pattern emerged for particular occupational groups (i.e. codes). There were similar numbers of discrepancies (when ordered ascending by the first coder) for each major group. For first coders with the first digit of '0' there was 0.8% of discrepant data with the second coder, '1'=18%, '2'=18%, '3'=13%, '4'=6%, '5'=11%, '6'=0.8%, '7'=8%, '8'=5%, '9'(valid)=9% and '9' (PISA)=11%. Note that '0' was the only new group for ISCO-08.

The pattern of discrepancies seemed related more to the National Centre/country than to the ISCO code groupings as demonstrated by variation in the distribution of discrepant cases between countries. The discrepant data included cases from CHE (1.2%), COL (3.3%), DNK (17.4%), HRV (7.2%), ISL (2.1%), ISR (45.2%), MEX (12.4%), RUS (3.5%), SVN (0.2%), TUR (0.2%), URG (1.0%) and SWE (5.8%). ISR accounted for the largest proportion (almost half) of discrepancies between coders.

For comparison, 2009 data were extracted to examine if the amount of discrepancy in the FT12 is comparable to when coders where using ISCO-88 classification. Five of the countries that participated in the FT12 double coding exercise also participated in the Main Survey 2009; however difficulties were encountered when attempting to access the original data for Chinese Taipei which left Chile, Denmark, Iceland and Portugal for the analyses.

It should be noted that in 2009 the number of double coded responses was more than 32,000 cases from four countries, namely Chile (11 178), Denmark (7 869), Iceland (435) and Portugal (12 572). As for the FT12 data, only those cases that had discrepant data between the two coders were selected. This consisted of 6198 cases (8% of mother's occupation codes and 11% of father's occupation codes with missing/invalid data removed). Filtering out those cases that matched on the first two digits, only 5% of mother and 7% of father occupation codes differed.

Double coding of occupational data appears to reveal a high level of consistency between coders. The data from the FT12 and Main Survey 2009 (MS09) suggest that ISCO-08 could be contributing to an improved accuracy. There did not appear to be any obvious issue with the coding of particular occupational groups but issues may be more associated with local training or understanding of ISCO coding in general – but this does not appear to be specific to ISCO-08 and perhaps equally existed in previous cycles of PISA. Considering the small proportion of discrepant data, it is suggested that the change in classification will not affect the integrity of the ISCO data and may even improve it.

QUESTIONNAIRE ROTATION AND IMPLICATIONS FOR THE INTERNATIONAL DATABASE AND DATA ANALYSIS

The structure of the international database for the data obtained from the student questionnaire is as follows. Questions in the common part contain responses from all students.

For questions in the rotated part, responses from two thirds of students are available. For each student, the international database contains data on all variables whereby missing are recorded with a '7' for those questions that were not administered to a student to indicate that this missing is a result of the rotation. As a consequence, each student will have all variables coded: the ones that s/he has answered reflecting his or her responses and the ones that s/he was not administered showing a '7'. So, in the rotated part, each student record contains actual responses to two thirds of the questions while one third is shown as '7', not administered due to the rotation design. Consistent with their use in the whole database, the non-response values (8) and (9), can, of course, still occur to indicate invalid (8) or missing (9) values where students left a question that they were administered unanswered.

For illustrative purposes a screenshot of selected variables from the student questionnaire file of the PISA 2012 international database is presented in Table 17.6. As it would have been too large to show all Student Questionnaire questions, this particular screenshot was taken of selected variables from the common and rotated part of the Student Questionnaire using actual data for the first twenty students in the data file for Albania. The variables in the file are as follows:

- CNT, SCHOOLID, STIDStd: Country code, 3 digit character; School ID, 7 digit; Student ID;
- ST01Q01 and ST04Q01: Questions from the common part (i.e. "International Grade" and "Gender"); hence valid codes for all cases;

- ST29Q01 and ST35Q01: Questions from the question sets included in Forms A and B (i.e. "Maths Interest - Enjoy Reading" and "Subjective Norms - Friends Do Well in Mathematics"); hence the cases with QUESTID=3 show the value of "7";
- ST53Q01 and ST55Q01: Questions from the question sets included in Forms A and C (i.e. " Learning Strategies - Important Parts vs. Existing Knowledge vs. Learn by Heart" and " Out of school lessons - <test lang>"); hence the cases with QUESTID=2 show the value of "7";
- ST77Q01 and ST79Q01: Questions from the question sets in Forms B and C (i.e."Math Teaching - Teacher Shows Interest" and "Teacher-Directed Instruction - Sets Clear Goals"); hence the cases with QUESTID=1 show the value of "7"; and
- QUESTID: Indicates which Student Questionnaire Form is used: 1=Form A, 2=Form B, 3=Form C; plus 5=Form UH (one hour, see Chapter 2 for more details) which contained the common part only (see Chapter 3 for more details on Student Questionnaire Forms. The questionnaire forms are available on *www.oecd.org/pisa*).

Therefore, going from right to left in Table 17.6 below, the data show that students 2 and 3 in the first school were administered Student Questionnaire Form A as indicated by the "1" for QUESTID in the last column. The two columns preceding this last column contain information on the variables ST77Q01 ("Math Teaching - Teacher Shows Interest") and ST79Q01 ("Teacher-Directed Instruction - Sets Clear Goals"). As these variables were in question set 3 which was not included in Student Questionnaire Form A, the code "7" indicates missing data due to these questions not having been administered to students 2 and 3.

Table 17.6 Selected variables in the student questionnaire file of the PISA 2012 database

CNT	SCHOOLID	STIDStd	ST01Q01	ST04Q01	ST29Q01	ST35Q01	ST53Q01	ST55Q01	ST77Q01	ST79Q01	QUESTID
ALB	0000001	00001	10	1	2	3	7	7	1	4	2
ALB	0000001	00002	10	1	3	1	2	1	7	7	1
ALB	0000001	00003	9	1	2	1	1	2	7	7	1
ALB	0000001	00004	9	1	7	7	2	1	1	9	3
ALB	0000001	00005	9	1	3	1	7	7	1	3	2
ALB	0000001	00006	9	1	7	7	2	1	1	1	3
ALB	0000001	00007	10	1	7	7	3	2	1	1	3
ALB	0000001	00008	10	2	2	2	7	7	1	1	2
ALB	0000001	00009	9	1	2	2	7	7	4	1	2
ALB	0000001	00010	10	1	7	7	2	1	2	3	3
ALB	0000002	00011	10	1	7	7	2	1	1	2	3
ALB	0000002	00012	10	1	2	2	2	1	7	7	1
ALB	0000002	00013	10	1	1	1	7	7	9	3	2
ALB	0000002	00014	10	1	2	1	1	4	7	7	1
ALB	0000002	00015	10	2	2	2	1	3	7	7	1
ALB	0000002	00016	10	2	1	2	7	7	1	1	2
ALB	0000002	00017	10	1	1	1	7	7	1	1	2
ALB	0000002	00018	10	2	3	9	3	1	7	7	1
ALB	0000002	00019	10	1	7	7	2	1	1	1	3
ALB	0000002	00020	10	1	7	7	1	2	1	1	3

The next four columns to the left show valid codes for students 2 and 3 who responded to variables in question sets 1 and 2 which were included in Form A, namely ST29Q01, ST35Q01, ST53Q01 and ST55Q01. In contrast, student 1, who responded to Student Questionnaire Form B as indicated by the "2" in QUESTID, has missing data due to the non-administration (i.e. code "7") recorded for the variables ST53Q01 and ST55Q01. Further, student 4, who responded to Student Questionnaire Form C as indicated by the "3" in QUESTID, has missing data due to the non-administration (i.e. code "7") recorded for the variables ST29Q01 and ST35Q01. It should also be noted that information on ST79Q01 is missing for student 4 despite the fact that this question was administered to this student. This is indicated by the code "9".

The two variables ST01Q01 and ST04Q01 contain valid codes for all students as the corresponding questions were included in the common part of the student questionnaire. While ST01Q01 shows that students 1 and 2 are in grade 10 and students 3 and 4 are in grade 9 all four students are female as indicated by the "1" in ST04Q01.

Implications for further analyses

The implications of the rotated Student Questionnaire design for further analyses differ depending on:

a. Whether the variables are located in the common or rotated part;

b. The intention of the analysis (e.g. examination of effects of school level variables on student performance).

It should be noted that rotation has no implication where proportions are reported. It also has no implications on the computation of standard errors or the use of replicate weights which will still provide the correct estimates. Likewise, rotation has no implications for the way that the replicate weights need to be used. The sum of the weights will be an estimate of the population size for students who responded to these items.

Rotation does have implications where frequencies or raw counts are reported and these values are used to estimate the population value of a particular variable. In this case, to get population estimates, raw values have to be multiplied by the inverse of the proportion who responded to the question. For questions from one question set to which 2/3 of students responded, values would have to be multiplied by a factor of 1.5.

Questions in the common part

Analyses using data from questions in the common part (e.g. the component questions constituting ESCS as well as gender, immigrant status and language at home) will be the same as in previous cycles as all students responded to these questions.

Details regarding the questions in the common part are given in Chapter 3.

Questions in the rotated part

While details regarding the questions in the rotated part of the Student Questionnaire are given in Chapter 3, this section provides further information regarding the implications of the rotated design on a) the set up of the international database, b) standard errors, c) effective sample sizes, d) intraclass correlations (rhos) of constructs and e) correlations between constructs in the Student Questionnaire and mathematics performance.

Due to the design of the Student Questionnaire, variables in the rotated part contain responses from two thirds of students in the sample. Thus, if a country has an average sample size of 30 students within a school, responses are obtained from 20 students. As an example, the analysis could include the constructs of Instrumental Motivation and Mathematics Interest (ST29Q01) and Subjective Norms (ST35Q01).

It should be noted that accuracy has been reduced slightly for variables in the rotated parts as the reduced sample size has increased the standard errors of the population estimates of the rotated variables. The increase in standard error for countries with a low intraclass correlation will be greater compared to countries with a high intraclass correlation. The range of the increase in standard errors of means is given in Table 17.7 while the effective sample sizes for different rotation designs are presented in Table 17.8.

Table 17.7 Range of the increase of the standard error of the means

	intraclass correlation, rho									
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Rotated Student Questionnaire design PISA 2012	1.22	1.05	1.03	1.02	1.01	1.01	1.00	1.00	1.00	1.00

Table 17.8 Effective sample sizes for different rotation designs

	intraclass correlation, rho									
	0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
No rotation	5250	1193	673	469	360	292	245	212	186	166
Rotated Student Questionnaire design PISA 2012	3500	1082	640	455	352	288	243	210	186	166

As can be seen in Table 17.7, rotation has little impact on standard errors for higher rhos. Correspondingly, Table 17.8 shows that the impact of the rotation in terms of decrease to effective sample size declines as rho increases.

Intraclass correlations for mathematics achievement (MAch) and all questionnaire indices for OECD countries are included in Table 17.9 for PISA 2003 and Table 17.10 for PISA 2012. As can be seen, the average intraclass correlation for questionnaire scales across OECD countries in PISA 2003 was 0.07 which meant that the increase in the standard error of the means for these scales was between 1.22 and 1.05 (see Table 17.7). Correspondingly for PISA 2012, the average intraclass correlation for questionnaire scales across OECD countries was 0.10 which meant an average increase of 1.05 in the standard error of the means for these scales (Table 17.7).

[Part 1/2]
Table 17.9 Intraclass correlation for questionnaire constructs in the student questionnaire for OECD countries in PISA 2003

CNT	MAch	ANXMAT	ATSCHL	ATTCOMP	BELONG	COMPHOME	COMPLRN	COOPLRN	CSTRAT	CULTPOS	DISCLIM	ELAB	ESCS	HEDRES
AUS	0.21	0.03	0.04	0.03	0.03	0.07	0.04	0.02	0.03	0.11	0.08	0.02	0.26	0.05
AUT	0.55	0.06	0.03	0.05	0.03	0.12	0.08	0.03	0.02	0.17	0.17	0.07	0.32	0.07
BEL	0.56	0.03	0.05	0.03	0.03	0.21	0.05	0.02	0.07	0.14	0.12	0.06	0.32	0.17
CAN	0.17	0.03	0.05	0.03	0.03	0.09	0.04	0.03	0.05	0.09	0.09	0.03	0.19	0.04
CHE	0.33	0.04	0.04	0.03	0.05	0.06	0.06	0.04	0.03	0.08	0.11	0.03	0.19	0.07
CZE	0.52	0.07	0.05	0.05	0.05	0.11	0.05	0.04	0.04	0.12	0.20	0.04	0.30	0.09
DEU	0.58	0.02	0.03	0.02	0.01	0.09	0.06	0.02	0.02	0.12	0.11	0.04	0.31	0.08
DNK	0.13	0.04	0.04	0.02	0.03	0.05	0.04	0.03	0.03	0.11	0.11	0.02	0.19	0.06
ESP	0.20	0.04	0.05	n/a	0.05	0.11	0.04	0.03	0.03	0.10	0.13	0.03	0.25	0.04
FIN	0.05	0.02	0.02	0.02	0.02	0.03	0.01	0.02	0.01	0.05	0.10	0.00	0.11	0.02
FRA	0.46	0.02	0.04	n/a	0.01	0.12	0.05	0.03	0.03	0.15	0.10	0.04	0.29	0.07
GBR	0.23	0.04	0.04	0.05	0.03	0.06	0.03	0.04	0.04	0.13	0.09	0.03	0.19	0.05
GRC	0.36	0.04	0.06	0.00	0.02	0.14	0.01	0.01	0.02	0.16	0.08	0.01	0.29	0.18
HUN	0.59	0.06	0.05	0.06	0.04	0.23	0.04	0.01	0.01	0.27	0.19	0.03	0.44	0.20
IRL	0.17	0.03	0.03	0.00	0.00	0.08	0.04	0.02	0.02	0.07	0.08	0.02	0.21	0.05
ISL	0.04	0.01	0.04	0.02	0.01	0.02	0.01	0.03	0.02	0.04	0.22	0.02	0.17	0.01
ITA	0.53	0.03	0.06	0.07	0.04	0.14	0.10	0.05	0.07	0.15	0.18	0.08	0.30	0.07
JPN	0.54	0.03	0.01	0.04	0.04	0.11	0.05	0.04	0.04	0.15	0.27	0.03	0.27	0.14
KOR	0.42	0.02	0.02	0.02	0.04	0.08	0.14	0.08	0.14	0.13	0.09	0.06	0.30	0.12
LUX	0.32	0.02	0.06	n/a	0.03	0.07	0.02	0.04	0.02	0.11	0.04	0.04	0.24	0.02
MEX	0.39	0.05	0.08	0.05	0.07	0.30	0.04	0.05	0.06	0.17	0.10	0.06	0.34	0.15
NLD	0.63	0.04	0.05	n/a	0.04	0.10	0.05	0.03	0.04	0.12	0.10	0.05	0.23	0.12
NOR	0.07	0.03	0.05	n/a	0.01	0.02	0.02	0.03	0.01	0.06	0.09	0.02	0.12	0.02
NZL	0.18	0.04	0.03	0.04	0.01	0.09	0.06	0.01	0.03	0.06	0.06	0.05	0.17	0.06
POL	0.13	0.02	0.03	0.01	0.02	0.11	0.02	0.02	0.01	0.09	0.11	0.02	0.23	0.05
PRT	0.34	0.03	0.03	0.02	0.03	0.16	0.01	0.03	0.05	0.11	0.08	0.02	0.24	0.09
SVK	0.43	0.06	0.05	0.06	0.04	0.17	0.06	0.03	0.04	0.10	0.15	0.05	0.32	0.19
SWE	0.11	0.03	0.03	0.03	0.01	0.03	0.04	0.01	0.02	0.08	0.11	0.02	0.12	0.02
TUR	0.56	0.07	0.05	0.03	0.03	0.24	0.04	0.03	0.05	0.14	0.10	0.04	0.37	0.24
USA	0.26	0.03	0.03	0.01	n/a	0.12	0.04	0.03	0.03	0.10	0.08	0.05	0.23	0.06

Note: n/a indicates country not administering questions forming construct.

[Part 2/2]
Table 17.9 Intraclass correlation for questionnaire constructs in the student questionnaire for OECD countries in PISA 2003

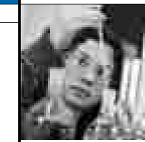
CNT	HIGHCONF	HOMEPOS	INSTMOT	INTCONF	INTMAT	INTUSE	MATHEFF	MEMOR	PRGUSE	ROUTCONF	SCMAT	STUREL	TEACHSUP	AVG*
AUS	0.03	0.17	0.02	0.05	0.04	0.03	0.08	0.02	0.03	0.02	0.03	0.06	0.05	0.05
AUT	0.09	0.22	0.17	0.11	0.09	0.08	0.19	0.03	0.10	0.15	0.06	0.10	0.14	0.11
BEL	0.04	0.24	0.06	0.13	0.05	0.05	0.14	0.03	0.07	0.11	0.02	0.07	0.07	0.09
CAN	0.05	0.12	0.05	0.06	0.05	0.05	0.06	0.03	0.09	0.05	0.04	0.07	0.06	0.06
CHE	0.04	0.12	0.07	0.05	0.03	0.03	0.12	0.04	0.04	0.06	0.03	0.13	0.09	0.06
CZE	0.10	0.18	0.10	0.09	0.06	0.06	0.22	0.06	0.07	0.12	0.05	0.08	0.15	0.10
DEU	0.03	0.20	0.04	0.04	0.04	0.02	0.12	0.04	0.04	0.06	0.01	0.07	0.09	0.07
DNK	0.02	0.13	0.03	0.03	0.05	0.03	0.05	0.02	0.03	0.03	0.03	0.12	0.08	0.06
ESP	n/a	0.15	0.04	n/a	0.04	n/a	0.07	0.02	n/a	n/a	0.05	0.10	0.11	0.07
FIN	0.02	0.05	0.01	0.01	0.02	0.01	0.03	0.01	0.02	0.01	0.02	0.05	0.07	0.03
FRA	n/a	0.23	0.03	n/a	0.04	n/a	0.15	0.03	n/a	n/a	0.05	0.07	0.08	0.08
GBR	0.11	0.16	0.04	0.07	0.04	0.07	0.09	0.05	0.14	0.09	0.04	0.06	0.06	0.07
GRC	0.06	0.28	0.03	0.06	0.03	0.04	0.10	0.01	0.04	0.09	0.05	0.10	0.08	0.08
HUN	0.09	0.37	0.04	0.16	0.05	0.07	0.22	0.02	0.07	0.21	0.05	0.11	0.09	0.12
IRL	0.02	0.13	0.03	0.05	0.02	0.04	0.07	0.01	0.02	0.03	0.03	0.04	0.07	0.05
ISL	0.03	0.04	0.03	0.03	0.03	0.03	0.04	0.01	0.02	0.02	0.02	0.10	0.12	0.04
ITA	0.06	0.21	0.11	0.10	0.11	0.06	0.15	0.04	0.08	0.12	0.05	0.11	0.13	0.10
JPN	0.05	0.19	0.09	0.07	0.06	0.04	0.26	0.03	0.11	0.09	0.03	0.07	0.08	0.09
KOR	0.02	0.21	0.09	0.03	0.08	0.05	0.21	0.05	0.03	0.06	0.10	0.06	0.04	0.09
LUX	n/a	0.16	0.04	n/a	0.03	n/a	0.08	0.02	n/a	n/a	0.01	0.07	0.07	0.06
MEX	0.17	0.28	0.05	0.28	0.10	0.14	0.07	0.06	0.08	0.28	0.06	0.09	0.09	0.12
NLD	n/a	0.18	0.03	n/a	0.04	n/a	0.09	0.04	n/a	n/a	0.04	0.07	0.06	0.07
NOR	n/a	0.06	0.03	n/a	0.03	n/a	0.04	0.01	n/a	n/a	0.02	0.08	0.07	0.04
NZL	0.04	0.14	0.02	0.02	0.07	0.05	0.06	0.03	0.06	0.02	0.02	0.05	0.04	0.05
POL	0.05	0.15	0.03	0.11	0.03	0.10	0.05	0.00	0.04	0.08	0.02	0.06	0.07	0.06
PRT	0.03	0.21	0.03	0.09	0.03	0.06	0.11	0.03	0.02	0.10	0.03	0.06	0.07	0.07
SVK	0.10	0.23	0.10	0.16	0.07	0.12	0.23	0.03	0.08	0.17	0.06	0.14	0.17	0.11
SWE	0.03	0.07	0.02	0.06	0.04	0.04	0.06	0.01	0.02	0.03	0.03	0.06	0.04	0.04
TUR	0.04	0.36	0.04	0.15	0.07	0.07	0.20	0.03	0.08	0.15	0.07	0.05	0.05	0.11
USA	0.04	0.17	0.03	0.07	0.05	0.04	0.06	0.04	0.05	0.06	0.04	0.05	0.06	0.06

*Average rho across all questionnaire scales in this country.
Note: n/a indicates country not administering questions forming construct.

[Part 1/4]
Table 17.10 Intraclass correlation for questionnaire constructs in the student questionnaire for OECD countries in PISA 2012

CNT	Math	ANXMAT	ATSCHL	ATTLNACT	BELONG	CLSMAN	COGACT	CULTDIST	CULTPOS	DISCLIMA	ENTUSE	ESCS	EXAPPLM	EXPUREM
AUS	0.28	0.04	0.08	0.05	0.07	0.09	0.06	n/a	0.10	0.11	0.05	0.24	0.05	0.09
AUT	0.54	0.06	0.10	0.07	0.10	0.11	0.05	n/a	0.16	0.14	0.05	0.33	0.02	0.20
BEL	0.55	0.06	0.04	0.05	0.05	0.11	0.07	0.08	0.12	0.11	0.02	0.29	0.03	0.19
CAN	0.20	0.04	0.05	0.04	0.04	0.08	0.05	n/a	0.08	0.08	n/a	0.16	0.04	0.07
CHE	0.34	0.06	0.06	0.06	0.09	0.09	0.04	n/a	0.09	0.10	0.04	0.17	0.06	0.13
CHL	0.55	0.04	0.11	0.03	0.04	0.18	0.08	n/a	0.13	0.15	0.12	0.64	0.04	0.09
CZE	0.56	0.06	0.07	0.06	0.05	0.24	0.11	n/a	0.12	0.26	0.07	0.27	0.11	0.19
DEU	0.55	0.03	0.04	0.04	0.05	0.05	0.03	n/a	0.10	0.07	0.03	0.27	0.01	0.16
DNK	0.18	0.04	0.08	0.03	0.03	0.14	0.08	0.45	0.09	0.13	0.03	0.18	0.02	0.03
ESP	0.17	0.03	0.05	0.05	0.06	0.10	0.07	n/a	0.09	0.11	0.03	0.22	0.05	0.06
EST	0.19	0.03	0.08	0.03	0.06	0.21	0.11	n/a	0.08	0.17	0.02	0.23	0.08	0.06
FIN	0.13	0.02	0.04	0.03	0.05	0.07	0.05	0.41	0.04	0.08	0.02	0.12	0.03	0.04
FRA	0.58	0.03	0.03	0.03	0.04	0.10	0.04	n/a	0.20	0.11	n/a	0.28	0.03	0.13
GBR	0.24	0.05	0.06	0.05	0.05	0.08	0.05	n/a	0.08	0.08	n/a	0.17	0.05	0.06
GRC	0.39	0.05	0.05	0.02	0.02	0.12	0.04	n/a	0.13	0.12	0.02	0.31	0.03	0.09
HUN	0.65	0.10	0.10	0.05	0.06	0.15	0.10	n/a	0.28	0.24	0.06	0.40	0.03	0.19
IRL	0.19	0.03	0.03	0.01	0.02	0.08	0.02	n/a	0.06	0.10	0.03	0.21	0.03	0.05
ISL	0.13	0.02	0.07	0.04	0.04	0.20	0.07	n/a	0.07	0.18	0.00	0.15	0.04	0.01
ISR	0.43	0.03	0.06	0.08	0.04	0.05	0.03	n/a	0.11	0.08	0.06	0.27	0.04	0.08
ITA	0.53	0.05	0.07	0.05	0.06	0.11	0.10	0.23	0.15	0.16	0.05	0.24	0.05	0.17
JPN	0.54	0.03	0.04	0.04	0.04	0.09	0.10	n/a	0.13	0.19	0.02	0.23	0.07	0.16
KOR	0.39	0.02	0.09	0.07	0.06	0.10	0.09	0.03	0.10	0.17	0.02	0.22	0.06	0.20
LUX	0.33	0.03	0.03	0.02	0.06	0.03	0.02	n/a	0.13	0.04	n/a	0.28	0.00	0.04
MEX	0.36	0.04	0.07	0.05	0.06	0.10	0.08	n/a	0.14	0.11	0.28	0.47	0.04	0.07
NLD	0.67	0.03	0.05	0.03	0.05	0.10	0.07	n/a	0.08	0.09	0.01	0.18	0.02	0.32
NOR	0.15	0.02	0.04	0.01	0.03	0.09	0.06	n/a	0.07	0.13	0.00	0.09	0.05	0.05
NZL	0.25	0.06	0.04	0.02	0.01	0.05	0.04	n/a	0.07	0.09	0.04	0.24	0.02	0.06
POL	0.25	0.04	0.06	0.03	0.05	0.17	0.07	n/a	0.08	0.18	0.02	0.28	0.03	0.05
PRT	0.32	0.02	0.05	0.05	0.04	0.12	0.05	0.06	0.13	0.09	0.01	0.30	0.02	0.05
SVK	0.46	0.07	0.07	0.06	0.07	0.18	0.10	0.02	0.13	0.17	0.12	0.37	0.07	0.10
SVN	0.59	0.05	0.14	0.09	0.08	0.16	0.07	0.09	0.14	0.22	0.05	0.25	0.05	0.14
SWE	0.15	0.02	0.05	0.02	0.03	0.10	0.07	n/a	0.08	0.11	0.01	0.14	0.03	0.06
TUR	0.62	0.04	0.06	0.03	0.05	0.10	0.04	n/a	0.11	0.10	0.11	0.30	0.02	0.11
USA	0.25	0.02	0.04	0.02	0.02	0.05	0.02	n/a	0.06	0.08	n/a	0.28	0.01	0.04

Note: n/a indicates country not administering questions forming construct.



[Part 2/4]
Table 17.10 Intraclass correlation for questionnaire constructs in the student questionnaire for OECD countries in PISA 2012

CNT	FAILMAT	FAMCON	FAMCONC	HEDRES	HERITCUL	HISEI	HOMEPOS	HOMSCH	HOSTCUL	ICTATTNEG	ICTATTPOS	ICTHOME	ICTSCH
AUS	0.05	0.14	0.13	0.09	n/a	0.14	0.17	0.20	n/a	0.04	0.06	0.03	0.14
AUT	0.02	0.32	0.32	0.12	n/a	0.27	0.17	0.23	n/a	0.06	0.08	0.05	0.14
BEL	0.10	0.21	0.23	0.17	0.07	0.24	0.19	0.13	0.09	0.02	0.06	0.02	0.24
CAN	0.05	0.07	0.10	0.07	n/a	0.10	0.13	n/a	n/a	n/a	n/a	n/a	n/a
CHE	0.05	0.25	0.19	0.05	n/a	0.14	0.10	0.11	n/a	0.03	0.05	0.03	0.27
CHL	0.04	0.29	0.21	0.18	n/a	0.52	0.56	0.11	n/a	0.04	0.03	0.26	0.11
CZE	0.06	0.23	0.19	0.13	n/a	0.24	0.14	0.18	n/a	0.08	0.05	0.04	0.14
DEU	0.04	0.21	0.29	0.11	n/a	0.26	0.12	0.11	n/a	0.03	0.03	0.03	0.20
DNK	0.04	0.09	0.06	0.09	0.26	0.12	0.13	0.14	0.41	0.04	0.03	0.03	0.09
ESP	0.04	0.09	0.15	0.04	n/a	0.18	0.12	0.09	n/a	0.09	0.03	0.04	0.15
EST	0.08	0.10	0.08	0.06	n/a	0.18	0.12	0.11	n/a	0.06	0.04	0.04	0.11
FIN	0.03	0.12	0.05	0.05	0.56	0.09	0.06	0.08	0.44	0.02	0.03	0.02	0.14
FRA	0.02	0.19	0.24	0.07	n/a	0.22	0.19	n/a	n/a	n/a	n/a	n/a	n/a
GBR	0.03	0.12	0.09	0.06	n/a	0.12	0.12	n/a	n/a	n/a	n/a	n/a	n/a
GRC	0.02	0.08	0.12	0.10	n/a	0.24	0.21	0.06	n/a	0.03	0.02	0.07	0.09
HUN	0.05	0.23	0.31	0.19	n/a	0.30	0.31	0.04	n/a	0.11	0.02	0.09	0.06
IRL	0.02	0.06	0.06	0.05	n/a	0.13	0.15	0.06	n/a	0.01	0.02	0.03	0.11
ISL	0.02	0.05	0.06	0.05	n/a	0.13	0.07	0.09	n/a	0.01	0.05	0.02	0.19
ISR	0.01	0.16	0.11	0.09	n/a	0.25	0.21	0.19	n/a	0.09	0.08	0.08	0.15
ITA	0.04	0.25	0.25	0.07	0.18	0.21	0.16	0.06	0.20	0.06	0.04	0.04	0.16
JPN	0.02	0.25	0.19	0.11	n/a	0.12	0.15	0.08	n/a	0.01	0.03	0.07	0.13
KOR	0.01	0.35	0.24	0.11	0.16	0.12	0.15	0.11	0.09	0.01	0.05	0.07	0.07
LUX	0.01	0.10	0.11	0.04	n/a	0.31	0.11	n/a	n/a	n/a	n/a	n/a	n/a
MEX	0.05	0.17	0.08	0.21	n/a	0.31	0.43	0.33	n/a	0.04	0.07	0.33	0.22
NLD	0.02	0.20	0.19	0.08	n/a	0.15	0.10	0.11	n/a	0.06	0.02	0.02	0.13
NOR	0.02	n/a	n/a	0.04	n/a	0.07	0.07	0.15	n/a	0.01	0.03	0.02	0.14
NZL	0.02	0.14	0.08	0.09	n/a	0.17	0.19	0.13	n/a	0.05	0.02	0.02	0.13
POL	0.05	0.08	0.07	0.02	n/a	0.25	0.14	0.05	n/a	0.03	0.01	0.04	0.13
PRT	0.04	0.08	0.10	0.08	0.09	0.27	0.17	0.03	0.04	0.06	0.01	0.03	0.10
SVK	0.03	0.19	0.17	0.38	0.07	0.25	0.37	0.17	0.12	0.07	0.03	0.13	0.14
SVN	0.03	0.19	0.21	0.07	0.11	0.22	0.14	0.07	0.14	0.05	0.04	0.06	0.10
SWE	0.02	0.05	0.06	0.05	n/a	0.12	0.08	0.15	n/a	0.02	0.03	0.02	0.17
TUR	0.03	0.15	0.22	0.25	n/a	0.16	0.35	0.05	n/a	0.07	0.02	0.26	0.09
USA	0.02	0.09	0.12	0.10	n/a	0.17	0.19	n/a	n/a	n/a	n/a	n/a	n/a

Note: n/a indicates country not administering questions forming construct.

[Part 3/4]
Table 17.10 Intraclass correlation for questionnaire constructs in the student questionnaire for OECD countries in PISA 2012

CNT	INFOCAR	INFOJOB1	INFOJOB2	INSTMOT	INTMAT	MATBEH	MATHEFF	MATINTFC	MATWKETH	MTSUP	OPENPS	OUTHOURS	PARED	PERSEV
AUS	0.10	0.12	0.08	0.03	0.05	0.07	0.13	0.04	0.06	0.07	0.05	0.08	0.14	0.04
AUT	0.12	0.04	0.00	0.13	0.07	0.05	0.19	0.14	0.07	0.07	0.04	0.07	0.22	0.03
BEL	0.06	0.04	0.03	0.06	0.05	0.06	0.12	0.06	0.04	0.07	0.04	0.07	0.16	0.04
CAN	0.08	0.17	0.10	0.04	0.04	0.07	0.08	0.04	0.05	0.06	0.04	0.09	0.09	0.05
CHE	n/a	n/a	n/a	0.07	0.04	0.04	0.13	0.07	0.05	0.06	0.03	0.05	0.10	0.04
CHL	n/a	n/a	n/a	0.03	0.04	0.03	0.07	0.04	0.04	0.10	0.03	0.05	0.45	0.03
CZE	n/a	n/a	n/a	0.09	0.09	0.10	0.16	0.12	0.07	0.18	0.07	0.07	0.16	0.04
DEU	n/a	n/a	n/a	0.01	0.03	0.03	0.12	0.03	0.01	0.09	0.02	0.05	0.17	0.01
DNK	0.11	0.06	0.03	0.03	0.04	0.04	0.08	0.05	0.05	0.08	0.06	0.06	0.10	0.04
ESP	n/a	n/a	n/a	0.03	0.04	0.04	0.06	0.04	0.04	0.06	0.03	0.03	0.16	0.05
EST	n/a	n/a	n/a	0.02	0.06	0.06	0.07	0.03	0.07	0.12	0.02	0.08	0.14	0.03
FIN	0.05	0.07	0.08	0.04	0.04	0.02	0.04	0.02	0.03	0.05	0.02	0.03	0.06	0.03
FRA	n/a	n/a	n/a	0.02	0.05	0.02	0.15	0.02	0.04	0.09	0.05	0.07	0.14	0.03
GBR	n/a	n/a	n/a	0.04	0.05	0.07	0.09	0.03	0.04	0.06	0.04	0.10	0.08	0.04
GRC	n/a	n/a	n/a	0.03	0.06	0.04	0.08	0.04	0.06	0.08	0.04	0.09	0.20	0.05
HUN	0.12	0.07	0.03	0.08	0.08	0.07	0.26	0.08	0.05	0.12	0.09	0.09	0.28	0.04
IRL	0.06	0.09	0.05	0.04	0.03	0.02	0.07	0.04	0.04	0.04	0.00	0.07	0.11	0.01
ISL	n/a	n/a	n/a	0.01	0.03	0.02	0.05	0.02	0.04	0.11	0.02	0.04	0.08	0.01
ISR	n/a	n/a	n/a	0.07	0.10	0.16	0.09	0.05	0.06	0.06	0.04	0.05	0.17	0.06
ITA	0.05	0.05	0.03	0.09	0.11	0.09	0.15	0.11	0.09	0.07	0.06	0.13	0.14	0.06
JPN	n/a	n/a	n/a	0.09	0.08	0.08	0.25	0.03	0.05	0.08	0.07	0.20	0.22	0.05
KOR	0.10	0.11	0.10	0.11	0.09	0.15	0.22	0.05	0.11	0.08	0.08	0.19	0.15	0.04
LUX	0.11	0.10	0.01	0.03	0.03	0.03	0.09	0.02	0.01	0.05	0.03	0.01	0.17	0.01
MEX	n/a	n/a	n/a	0.05	0.08	0.06	0.07	0.05	0.06	0.08	0.06	0.06	0.34	0.03
NLD	n/a	n/a	n/a	0.04	0.04	0.04	0.11	0.03	0.02	0.07	0.02	0.06	0.08	0.02
NOR	n/a	n/a	n/a	0.01	0.03	0.02	0.04	0.04	0.02	0.05	0.01	0.03	0.06	0.02
NZL	0.09	0.08	0.01	0.02	0.06	0.09	0.11	0.02	0.03	0.03	0.03	0.08	0.11	0.03
POL	n/a	n/a	n/a	0.04	0.04	0.07	0.09	0.02	0.05	0.11	0.02	0.03	0.21	0.03
PRT	0.06	0.06	0.07	0.05	0.05	0.03	0.14	0.03	0.05	0.05	0.04	0.04	0.22	0.04
SVK	0.10	0.04	0.03	0.07	0.10	0.09	0.16	0.09	0.06	0.09	0.04	0.07	0.22	0.05
SVN	0.07	0.08	0.09	0.10	0.08	0.06	0.15	0.13	0.05	0.08	0.04	0.26	0.17	0.03
SWE	n/a	n/a	n/a	0.03	0.02	0.02	0.07	0.02	0.03	0.05	0.04	0.05	0.06	0.01
TUR	n/a	n/a	n/a	0.04	0.05	0.07	0.18	0.02	0.05	0.04	0.04	0.05	0.19	0.02
USA	n/a	n/a	n/a	0.01	0.05	0.04	0.08	0.02	0.03	0.02	0.02	0.07	0.22	0.02

Note: n/a indicates country not administering questions forming construct.

[Part 4/4]
Table 17.10 Intraclass correlation for questionnaire constructs in the student questionnaire for OECD countries in PISA 2012

CNT	SCMAT	SMINS	STUDREL	SUBNORM	TCHBEHFA	TCHBEHSO	TCHBEHTD	TEACHSUP	USEMATH	USESCH	WEALTH	AVG*
AUS	0.05	0.16	0.09	0.07	0.06	0.11	0.05	0.06	0.08	0.20	0.13	0.09
AUT	0.05	0.31	0.08	0.07	0.10	0.18	0.05	0.12	0.09	0.16	0.09	0.13
BEL	0.04	0.20	0.03	0.06	0.05	0.14	0.08	0.08	0.16	0.22	0.10	0.11
CAN	0.05	0.23	0.06	0.06	0.06	0.10	0.06	0.06	n/a	n/a	0.13	0.08
CHE	0.03	0.14	0.09	0.08	0.08	0.24	0.06	0.07	0.09	0.13	0.07	0.09
CHL	0.03	0.11	0.06	0.03	0.13	0.17	0.11	0.12	0.15	0.14	0.58	0.15
CZE	0.07	0.20	0.09	0.11	0.10	0.14	0.11	0.18	0.14	0.14	0.09	0.13
DEU	0.01	0.28	0.08	0.04	0.09	0.15	0.05	0.09	0.07	0.16	0.08	0.10
DNK	0.03	0.09	0.10	0.07	0.06	0.09	0.07	0.11	0.22	0.09	0.12	0.09
ESP	0.04	0.08	0.07	0.04	0.08	0.10	0.07	0.10	0.10	0.16	0.10	0.08
EST	0.04	0.11	0.08	0.03	0.10	0.18	0.16	0.12	0.10	0.12	0.08	0.09
FIN	0.02	0.11	0.05	0.06	0.05	0.08	0.06	0.07	0.10	0.10	0.05	0.08
FRA	0.03	0.12	0.04	0.04	0.08	0.12	0.08	0.10	n/a	n/a	0.10	0.10
GBR	0.05	0.18	0.06	0.06	0.06	0.12	0.07	0.06	n/a	n/a	0.09	0.08
GRC	0.04	0.13	0.07	0.06	0.09	0.17	0.06	0.10	0.09	0.09	0.15	0.09
HUN	0.07	0.40	0.07	0.13	0.15	0.25	0.13	0.18	0.14	0.12	0.15	0.15
IRL	0.02	0.11	0.04	0.03	0.06	0.10	0.05	0.05	0.08	0.11	0.10	0.06
ISL	0.02	0.17	0.07	0.05	0.04	0.09	0.12	0.13	0.12	0.19	0.06	0.07
ISR	0.02	0.09	0.09	0.09	0.15	0.15	0.06	0.07	0.20	0.12	0.30	0.11
ITA	0.08	0.29	0.08	0.08	0.11	0.15	0.10	0.12	0.15	0.13	0.10	0.12
JPN	0.03	0.60	0.06	0.10	0.05	0.11	0.08	0.08	0.02	0.12	0.10	0.12
KOR	0.07	0.73	0.08	0.14	0.05	0.06	0.05	0.05	0.05	0.17	0.08	0.12
LUX	0.01	0.27	0.03	0.03	0.05	0.11	0.03	0.06	n/a	n/a	0.05	0.07
MEX	0.04	0.13	0.06	0.06	0.11	0.13	0.10	0.12	0.08	0.16	0.45	0.14
NLD	0.03	0.04	0.04	0.04	0.08	0.16	0.07	0.07	0.10	0.11	0.05	0.09
NOR	0.02	0.11	0.05	0.04	0.07	0.08	0.07	0.08	0.12	0.14	0.05	0.06
NZL	0.02	0.16	0.06	0.07	0.05	0.10	0.05	0.04	0.08	0.13	0.14	0.08
POL	0.03	0.85	0.09	0.05	0.08	0.11	0.10	0.14	0.14	0.11	0.13	0.10
PRT	0.03	0.27	0.05	0.06	0.09	0.14	0.08	0.09	0.09	0.12	0.11	0.09
SVK	0.07	0.44	0.10	0.14	0.14	0.22	0.11	0.15	0.18	0.13	0.21	0.14
SVN	0.06	0.34	0.08	0.08	0.12	0.21	0.10	0.13	0.10	0.12	0.07	0.12
SWE	0.02	0.39	0.06	0.06	0.04	0.07	0.04	0.07	0.08	0.23	0.07	0.07
TUR	0.04	0.24	0.03	0.05	0.04	0.13	0.04	0.04	0.14	0.13	0.29	0.11
USA	0.03	0.11	0.06	0.03	0.05	0.12	0.04	0.03	n/a	n/a	0.19	0.07

*Average rho across all questionnaire scales in this country.
Note: n/a indicates country not administering questions forming construct.

Still, consideration has to be given to the possibilities of analysing responses to questions in the rotated part by subgroups of interest, such as immigrants or single-parents. Such analyses will have become more limited as group sizes decreased. In general, in PISA results are not reported if there are fewer than 30 students or fewer than five schools with valid data.

To explore the implications for analyses on correlations and corresponding standard errors, correlations between questionnaire indices and mathematics achievement computed using all student responses were compared to the correlation computed using 2/3 of the student responses. PISA 2003 data for Finland, Germany and Korea were used as these countries differ in terms of their intraclass correlation and geolocation. Thus, Germany has a high intraclass correlation with values differing across schools but tending to be similar for students in the same school whereas the reverse applies in Finland. Results of the analyses are shown in Table 17.11.

Results in Table 17.11 confirm earlier analyses (Adams, Lietz and Berezner, 2013; Berezner and Lietz, 2009) whereby the differences in correlation coefficients range from 0 to 0.02 and the corresponding standard errors differ from 0 to 0.01. Therefore, differences are considered negligible.

It should be noted that these comparative analyses could not be replicated using PISA 2012 data as, by design, information was available for only two thirds of students per question in the rotated parts of the student questionnaire in that cycle.

One alternative to assigning the missing data due to rotation a specific code (i.e. '7') would have been to attempt imputations as a means of replacing missing information with 'pseudo-information'. However, a number of considerations with respect to the optimal model, challenges from bidirectional imputation, logical inconsistencies in data patterns and precision led to the decision not to impute the values that were missing due to rotation in PISA 2012.

Table 17.11 Correlations between questionnaire indices and mathematics achievement (PISA 2003 data)

		All students		2/3 students		
CNT	**Index**	**Corr**	**SE**	**Corr**	**SE**	**Increase in SE (ratio: Col6/Col4)**
DEU	ANXMAT	-0.34	0.015	-0.35	0.018	1.20
DEU	ATSCHL	-0.10	0.019	-0.09	0.022	1.15
DEU	ATTCOMP	0.00	0.015	-0.02	0.022	1.49
DEU	BELONG	-0.02	0.018	-0.02	0.020	1.09
DEU	COMPHOME	0.26	0.016	0.25	0.022	1.39
DEU	COMPLRN	-0.06	0.018	-0.06	0.021	1.19
DEU	COOPLRN	-0.03	0.016	-0.04	0.019	1.14
DEU	CSTRAT	-0.08	0.020	-0.09	0.024	1.18
DEU	CULTPOS	0.25	0.014	0.27	0.018	1.27
DEU	DISCLIM	0.22	0.019	0.24	0.022	1.14
DEU	ELAB	-0.06	0.018	-0.05	0.022	1.24
DEU	ESCS	0.48	0.014	0.49	0.016	1.17
DEU	HEDRES	0.25	0.015	0.26	0.019	1.26
DEU	HIGHCONF	0.09	0.021	0.09	0.026	1.25
DEU	HOMEPOS	0.39	0.012	0.40	0.017	1.43
DEU	INSTMOT	0.02	0.019	0.03	0.023	1.20
DEU	INTCONF	0.21	0.017	0.21	0.020	1.22
DEU	INTMAT	0.12	0.019	0.14	0.020	1.10
DEU	INTUSE	0.05	0.019	0.03	0.025	1.33
DEU	MATHEFF	0.50	0.015	0.51	0.017	1.19
DEU	MEMOR	-0.22	0.019	-0.22	0.025	1.35
DEU	PRGUSE	-0.06	0.022	-0.06	0.027	1.25
DEU	ROUTCONF	0.32	0.015	0.32	0.016	1.04
DEU	SCMAT	0.27	0.016	0.29	0.018	1.14
DEU	STUREL	-0.07	0.021	-0.05	0.023	1.09
DEU	TEACHSUP	-0.12	0.020	-0.11	0.023	1.15
FIN	ANXMAT	-0.44	0.013	-0.44	0.015	1.16
FIN	ATSCHL	0.14	0.015	0.14	0.017	1.12
FIN	ATTCOMP	-0.01	0.013	-0.01	0.016	1.20
FIN	BELONG	-0.03	0.015	-0.03	0.018	1.21
FIN	COMPHOME	0.15	0.015	0.16	0.019	1.26
FIN	COMPLRN	0.23	0.015	0.24	0.019	1.29
FIN	COOPLRN	0.03	0.017	0.05	0.020	1.21
FIN	CSTRAT	0.11	0.013	0.13	0.019	1.53
FIN	CULTPOS	0.21	0.015	0.21	0.016	1.07
FIN	DISCLIM	0.12	0.015	0.12	0.020	1.33
FIN	ELAB	0.17	0.014	0.17	0.018	1.27
FIN	ESCS	0.33	0.014	0.32	0.016	1.13
FIN	HEDRES	0.18	0.017	0.18	0.020	1.20
FIN	HIGHCONF	0.09	0.015	0.09	0.019	1.27
FIN	HOMEPOS	0.27	0.015	0.27	0.018	1.15
FIN	INSTMOT	0.29	0.017	0.30	0.019	1.13
FIN	INTCONF	0.11	0.017	0.12	0.018	1.10
FIN	INTMAT	0.33	0.015	0.34	0.019	1.25
FIN	INTUSE	0.04	0.013	0.05	0.018	1.32
FIN	MATHEFF	0.52	0.013	0.51	0.015	1.21
FIN	MEMOR	0.08	0.017	0.08	0.023	1.38
FIN	PRGUSE	0.03	0.015	0.04	0.019	1.22
FIN	ROUTCONF	0.23	0.014	0.23	0.018	1.31
FIN	SCMAT	0.57	0.011	0.58	0.014	1.25
FIN	STUREL	0.10	0.014	0.11	0.018	1.26
FIN	TEACHSUP	0.05	0.018	0.04	0.021	1.18
KOR	ANXMAT	-0.22	0.014	-0.23	0.017	1.20
KOR	ATSCHL	0.00	0.016	0.01	0.016	0.99
KOR	ATTCOMP	0.02	0.018	0.03	0.021	1.13
KOR	BELONG	0.10	0.017	0.09	0.019	1.07
KOR	COMPHOME	0.26	0.015	0.26	0.019	1.26
KOR	COMPLRN	0.40	0.013	0.40	0.018	1.32
KOR	COOPLRN	0.34	0.014	0.34	0.016	1.16
KOR	CSTRAT	0.40	0.014	0.39	0.016	1.12
KOR	CULTPOS	0.27	0.022	0.26	0.025	1.14
KOR	DISCLIM	0.13	0.017	0.14	0.021	1.19
KOR	ELAB	0.30	0.014	0.31	0.018	1.29
KOR	ESCS	0.37	0.025	0.38	0.026	1.05
KOR	HEDRES	0.26	0.018	0.26	0.021	1.17
KOR	HIGHCONF	0.11	0.016	0.13	0.017	1.05
KOR	HOMEPOS	0.36	0.021	0.36	0.025	1.17
KOR	INSTMOT	0.35	0.015	0.35	0.016	1.10
KOR	INTCONF	0.20	0.016	0.19	0.021	1.30
KOR	INTMAT	0.39	0.014	0.40	0.016	1.13
KOR	INTUSE	-0.07	0.018	-0.08	0.021	1.16
KOR	MATHEFF	0.58	0.012	0.58	0.016	1.30
KOR	MEMOR	0.18	0.016	0.19	0.019	1.13
KOR	PRGUSE	0.08	0.016	0.10	0.021	1.30
KOR	ROUTCONF	0.32	0.014	0.33	0.016	1.11
KOR	SCMAT	0.46	0.014	0.47	0.016	1.16
KOR	STUREL	0.10	0.024	0.12	0.022	0.95
KOR	TEACHSUP	0.06	0.019	0.05	0.019	1.00

Imputations for missing data are model-dependent draws from the posterior distribution of random variables, conditional on the observed values of other available variables, using estimated relationships between the variable that is missing and the remainder of the variables. Many possible imputation models are conceivable which differ, among other things, in terms of the distributional assumptions that are made, the number of dimensions and the extent to which information from other observed student and school-level variables is used.

To do imputations properly for the whole of the international database would require an IRT (Item Response Theory) model with the number of dimensions equal to the reported proficiency scales plus the number of questionnaire scales with another dimension added for each missing background variable. The scope of these imputations would be well beyond what has been done to date.

The problem of bidirectional imputation would arise if the imputation for missing background data were to include proficiency data. Here, a problem of circularity may arise as plausible values for proficiencies are imputed using information from the background variables in the first place.

In the absence of imputed values for the student questionnaire data that are missing due to rotation, a few alternatives are recommended when using data from all three question sets. Central to the justification for these alternatives is that the missing data that results from rotation is missing completely at random.

First, since the design produces an estimate of the complete variance/covariance matrix modelling that utilises only the variance/covariance matrix can easily be undertaken.

Second, where possible full information maximum likelihood (FIML) can be implemented. Good introductions to FIML have been written by Marcoulides and Schumacker (1996) and Wothke (2000), among others. Software programs that allow FIML estimation include Amos and MPlus.

Third, analysts can impute data required for the specific models they seek to analyse. Software programs that enable imputation include MPlus, R, SAS®, SPSS® and Stata.

Intentions of the analyses

If the intended analyses of the PISA 2012 data are multilevel in nature, consideration needs to be given to any variables and constructs for which data were collected at the student level with the intention of using them as contextual variables at the school level. Thus, for example, individual students' ratings of their school's disciplinary climate are frequently aggregated and entered at the school level in multilevel models examining differences in student achievement. Lüdtke et al. (2009) have pointed out that the reliability of the observed group mean rating – labelled ICC(2) – is dependent on the proportion of the total variance that can be attributed to between-group differences – labelled ICC(1) – and the number of students from whom ratings have been obtained. The following equations define ICC(1) and ICC(2):

17.1

$$ICC(1) = \frac{MS_B - MS_W}{MS_B + (k-1)MS_W}$$

17.2

$$ICC(2) = \frac{k \times ICC(1)}{1 + (k-1) \times ICC(1)}$$

Where k is the number of students in each group; MS_B is the between mean square; and MS_w is the within mean square from analysis of variance.

In their article, Lüdtke et al. (2009) assumed a rather high ICC(1) of 0.43 indicating that 43% of the total variation found in all student ratings was associated with differences between groups. Combined with a group size of ten students, this translated to a reliability of the group mean rating ICC(2) of 0.88 which was a relatively high level of reliability for an attitudinal scale.

In Table 17.12, implications for the reliability of the group mean rating ICC(2) is given for a number of different combinations of ICC(1)s and numbers of students.

Table 17.12 Implications of different group sizes on reliabilities of group mean ratings

ICC(1)	ICC(2) for 20 students (= 3 rotated forms)	ICC(2) for 30 students = no rotation
0.43	0.94	0.96
0.30	0.90	0.93
0.25	0.87	0.91
0.20	0.83	0.88
0.15	0.78	0.84
0.10	0.69	0.77
0.05	0.51	0.61

Note: For definitions of ICC(1) and ICC(2) see equations 1 and 2 above.

Table 17.12 demonstrates that the rotation design has had different implications depending on the between-group differences associated with a particular scale or construct – which is likely to differ for the countries participating in PISA.

In such contextual analyses, Lüdtke et al. (2008) have recommended to use a multilevel latent covariate approach (MLC) instead of a multilevel manifest covariate (MMC) approach, particularly when a) the ICC(1) is small, b) the sampling proportion (i.e. in the PISA context the number of 15-year old students that the 30 randomly selected 15-year-olds in each school represent) is small and c) the number of cases are small. Through a number of simulation and real-data applications, the authors showed that the MLC could correct for the unreliability at level 2 for constructs that represented aggregates of student-level ratings. The authors particularly argued for the use of the MLC approach where the L2 construct was assumed to be a reflective aggregation of L1 measures. Reflective constructs, for example, would be individual students' ratings of instructional practices of the teacher (a common referent at L2). Formative L2 constructs for which no such adjustment might be necessary would include aggregations of L1 variables such as age, gender, or SES (socio-economic status).

In line with Lüdtke et al. (2009), it is recommended that future multilevel analyses of student-level ratings of instructional practices or school context at the between-school level should report both ICC(1) and ICC(2).

Analyses were conducted with data for five countries from PISA 2003, the cycle prior to 2012 in which mathematics had been the major domain (Van de gaer, Lietz and Adams, 2011). The countries, namely Australia, Finland, Germany, Korea and the United States were selected as they differed with respect to the following characteristics which Lüdtke et al. (2008) had specified as impacting on the group level effects:

- The number of level 1 cases per group (from 19 in the United States to 38 in Australia);
- The number of groups (from 149 in Korea to 321 in Australia);
- The ICC in the outcome variable, mathematics achievement (ranging from 0.06 in Finland to 0.63 in Germany).

In addition, these five countries differed in terms of their mathematics achievement in 2003 relative to the OECD average, with Finland, Korea and Australia performing above average, Germany around average and the United States below average. Finally, these countries also included some cultural variation.

The purpose of the analyses was to examine the differences between the multilevel latent covariate approach (MLC) and the multilevel manifest covariate approach (MMC) to modelling individual students' ratings of instructional practices or school context at the between-school level, particularly for reflective constructs where individual students' ratings related to a common referent at level 2. At the same time, the analyses explored the implications of the rotated student questionnaire design which resulted in having only two thirds of student responses available in the analyses. Results showed that the differences in estimates between the unrotated and the rotated questionnaire design were far smaller than the differences in estimates between the multilevel latent covariate approach (MLC) and the multilevel manifest covariate approach (MMC).

In addition, the results of the analyses seemed to confirm partially the findings of Lüdtke et al. (2008). Thus, the larger beta coefficients of the group-level effects using the multilevel latent covariate approach seemed to suggest that the unreliability of the group-level effects was taken into account. The fact that the SEs (standard errors) of the MMC approach in most cases were smaller than those of the MLC approach might have been expected because the group mean of the covariate was treated as observed whereas in the MLC it was not (leading to larger sampling variability).

Results showed the most pronounced difference in the beta coefficient of the group-level variable in the United States between the MLC and MMC approach. This could be expected as, according to Lüdtke et al. (2008), the difference

between the two approaches (MMC and MLC) were the most pronounced for small numbers of level 1 cases within each level 2 group and the United States had the lowest cluster size (=19) of the countries included in the analysis.

In other words, the MLC approach seemed to control for the unreliability in the group-level effects that was introduced when a reduced number of students were being sampled within schools. Using the MMC approach seemed to result in an underestimation of the group-level effect. Thus, the approaches MLC and MMC seemed to lead to different estimates of the group-level effect in multilevel analyses and more so for countries with low predictor ICC, relatively higher outcome ICC and relatively smaller average cluster size.

While differences were noted between the absolute sizes of the estimates for rotated and unrotated questionnaire designs as well as and the analytical approaches (i.e. MMC compared with MLC), the conclusions in terms of whether or not an effect would be considered significant remained the same in all instances.

Still, as Lüdtke et al. (2008) noted, most analysts use the MMC approach. Where analysts of PISA data have used this approach previously and want to replicate their analyses with PISA 2012 data, they should also use the MMC approach with the PISA 2012 data to ensure comparability of results. Alternatively, they may choose to rerun the earlier analysis using the MLC approach and also apply this approach to the PISA 2012 data.

Notes

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

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

2. This link is not available anymore, but for more details on the correspondance between ISCO-88 and ISCO-08 please refer to the following website: *http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm*.

Reference

Adams, R. J., P. Lietz and **A. Berezner** (2013), "On the use of rotated context questionnaires in conjunction with multilevel item response models", *Large-scale Assessments in Education*, No. 1(1).

Bartram, D. (2007), "Increasing validity with forced-choice criterion measurement formats", *International Journal of Selection and Assessment*, No. 15, pp. 263-272. *http://dx.doi.org/10.1111/j.1468-2389.2007.00386.x*

Bertling, J. P. and **P.C. Kyllonen** (2013), "Anchoring adjustments of student questionnaire indexes: possible scenarios for the PISA 2012 international database", Paper prepared for the PISA Technical Advisory Group (TAG) meeting, Melbourne, 4-5 April 2013.

Bertling, J. and **R. Roberts** (2011), "Student questionnaire (Opportunity to learn): Field Trial analyses", Paper presented at the QEG Meeting, Melbourne (QEG(1109)1b), Australia, 26-28 September 2011.

Berezner, A. and **P. Lietz** (2009), "Implication for Preferred Model of Rotation on Cognitive Proficiency Estimate Generation Reporting and Subsequent Analyses", Paper presented at the QEG Meeting, Offenbach, Germany, 19-21 October 2009.

Bowen, C.C., B.A. Martin and **S.T. Hunt** (2002), "A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires", International *Journal of Organizational Analysis*, No. 10(3), pp. 240-259.

Brown, A. and **A. Maydeu-Olivares** (2013), "How IRT can solve problems of ipsative data in forced-choice questionnaires", *Psychological Methods*, No. 18(1), p. 36.

Cheung, M.W.L. and **W. Chan** (2002), "Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis", *Structural Equation Modeling,* No. 9, pp. 55-77.

Christiansen, N., G. Burns and **G. Montgomery** (2005), "Reconsidering the use of forced-choice formats for applicant personality assessment", *Human Performance*, No. 18, pp. 267–307. *http://dx.doi.org/10.1207/s15327043hup1803_4*

Christian, M. S., B. D. Edwards and **J. C. Bradley**, (2010), "Situational judgment tests: Constructs assessed and a meta-analytic of their criterion-related validities", *Personnel Psychology*, No. 63, pp. 83-117.

De la Torre, J., V.Ponsoda, I. Leenen and **P. Hontangas** (2011), "Some extensions of the Multi-Unidimensional Pairwise Preference Model"*,* Paper presented at the 26th annual conference for the Society of Industrial and Organizational Psychology, Chicago.

Fischer, R., J. Fontaine, F. J. R. van de Vijver and **D. van Hemert** (2009), "What is style and what is bias in cross-cultural comparisons? An examination of response styles in cross-cultural research", in A. Gari and K. Mylonas (eds.), *Quod Erat Demonstrandum: From Herodotus' Ethnographic Journeys to Cross-Cultural Research*, Pedio, Athens, pp. 137 – 148.

Ganzeboom, Harry B.G. (2010), "Do's and Don't's of Occupation Coding with an Extension on ISCO-08", Working paper, Department of Social Research Methodology, VU Amsterdam, last revised January 3 2010.

Hopkins, D. J. and **G. King** (2010), "Improving anchoring vignettes designing surveys to correct interpersonal incomparability, *Public Opinion Quarterly*.

Kyllonen, P., J. Bertling and **R. Roberts** (2013), "Overclaiming: A new technique for correcting for response tendencies in international assessments", Paper presented at the 15th Biennial Conference EARLI 2013, Munich, Germany.

Kyllonen, P., P. Lietz and **R. Roberts** (2010), "Cross-cultural comparative questionnaire issues », Paper prepared for the PISA2012 Questionnaire Expert Group (QEG) meeting, Hong Kong, 25-27 February 2010.

Lievens, F., T. Buyse and **P. R. Sackett** (2005), "The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains", *Journal of Applied Psychology*, No. 90, pp. 442-452.

Lievens, F. and **P. Coestsier** (2002), "Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity", *International Journal of Selection and Assessment*, No.10, pp. 245-257.

Lüdtke, O., H. W. Marsh, A. Robitzsch, U.Trautwein, T. Asparouhov and **B. Muthén** (2008), "The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies", *Psychological Methods,* No. 13, pp. 203-229.

Lüdtke, O., A. Robitzsch, U. Trautwein and **M. Kunter** (2009), "Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modelling", *Contemporary Educational Psychology*, No. 34(2), pp. 120-131.

Maydeu-Olivares, A. and **A. Brown** (2010), "Item response modelling of paired comparison and ranking data", *Multivariate Behavioural Research,* No. 45, pp. 935-974.
http://dx.doi.org 10.1080/00273171.2010.531231

McDaniel, M.A., F.P. Morgeson, E.B. Finnegan, M.A. Campion and **E.P. Braverman** (2001), "Predicting job performance using situational judgment tests: A clarification of the literature", *Journal of Applied Psychology*, No. 86, pp. 730-740.

McDaniel, M. A., J. L Anderson., M. H. Derbish and **N. Morrisette** (2007), "Testing the testing effect in the classroom", *European Journal of Cognitive Psychology*, No. 19, pp. 494-513.

Marcoulides, G.A. and **R.E. Schumacker** (1996), *Advanced Structural Equation Modeling*, Erlbaum, Mahwah, New Jersey.

OECD (2013a), *PISA 2012 Results: Creative Problem Solving. Students' Skills in Tackling Real-Life Problems* (Volume V), PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264208070-en

OECD (2013b), *PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance to Succeed* (Volume II), PISA, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264201132-en

OECD (2009), *PISA 2006 Technical Report*, PISA, OECD Publishing, Paris, pp. 346-350.
http://dx.doi.org/10.1787/9789264048096-en

Paulhus, D. L., P. D. Harms, M. N. Bruce and **D. C. Lysy** (2003), "The over-claiming technique: Measuring self-enhancement independent of ability", *Journal of Personality and Social Psychology*, Vol. 84(4), April 2003, pp. 890-904. *http://dx.doi.org/10.1037/0022-3514.84.4.890*

Stark, S., O.S. Chernyshenko and **F. Drasgow** (2005), "An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: the multi-unidimensional pairwise-preference model", *Applied Psychological Measurement*, No. 29 (3), pp. 184-203.

Van de Gaer, E., P. Lietz and **R. Adams** (2011), "Application of the multivariate latent covariate model using PISA 2003", Paper presented at the QEG Meeting, Melbourne (QEG(1109)2b), Australia, 26-28 September 2011.

Van Soest, A., L. Delaney, C. Harmon, A. Kapteyn and **J. P. Smith** (2011), "Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions", *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, pp. 174(3), pp. 575-595.

Weekley J.A. and **R.E. Ployhart** (2006), *Situational judgment tests: Theory, management, and application*, Erlbaum, Mahwah, New Jersey.

White, L.A. and **M.C. Young** (1998), "Development and validation of the Assessment of Individual Motivation (AIM)", Paper presented at the Annual Meeting of the American Psychological Association, San Francisco.

Wothke, W. (2000), "Longitudinal and multi-group modeling with missing data", in T.D. Little, K.U. Schnabel, and J. Baumert (eds.), *Modeling Longitudinal and Multiple Group Data: Practical Issues, Applied Approaches and Specific Examples*, Erlbaum, Mahwah, New Jersey. (Reproduced with permission).

Ziegler, M., C. MacCann and **R.D. Roberts** (2012), *New Perspectives on Faking in Personality Assessment*, Cambridge University Press, Oxford.

Zimmerman, J., P.K.Broder, J.J. Shaughnessy and **B.J. Underwood** (1977), "A recognition test of vocabulary using signal-detection measures, and some correlates of word and nonword recognition", *Intelligence*, No. 1, pp. 5-31.