



3

Context Questionnaire Development

Introduction	48
A sustainable framework for the PISA context questionnaires.....	48
Cross-cultural comparability of measures in the context questionnaires	51
Transitioning the context questionnaires from paper administration to online administration mode	54
Update of the International Standard Classification of Occupations (ISCO) from its 1988 version to the 2008 version	55
The measurement of Opportunity to learn	56
The rotation of the Student Context Questionnaire – design and intended analyses	58



INTRODUCTION

The goal of the PISA background instruments is to gather data that can help policy makers and educators understand why and how students achieve certain levels of performance. PISA questionnaires must cover the most important antecedents and processes of student learning at the individual, school, and system level. The questionnaires also allow the collection of non-cognitive student performance outcomes such as student attitudes, interests, motivations, and beliefs.

At the same time, with the programme undertaking its fifth assessment, a number of points regarding the PISA context questionnaires required attention, including:

- Developing a sustainable framework for the context questionnaires that would ensure the monitoring of essential contextual characteristics over time while at the same time enabling new topics to be incorporated.
- Addressing questions regarding the cross-cultural comparability of measures in the context questionnaires.
- Transitioning the context questionnaires from paper administration to online administration mode.
- Updating the coding of parental occupation according to the 2008 International Standard Classification of Occupations (ISCO-08) from its 1988 version to the 2008 version.

In addition, the Consortium set itself the challenge of two further innovations in PISA 2012:

- the expanded measurement of Opportunity to Learn (OTL); and
- the rotation of the Student Context Questionnaire.

This chapter provides an overview of the questionnaires and their development process, while Chapter 16 describes questionnaire index construction and Chapter 17 describes the research that was undertaken during questionnaire construction and validation.

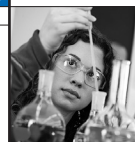
A SUSTAINABLE FRAMEWORK FOR THE PISA CONTEXT QUESTIONNAIRES

For PISA 2012, the conceptual framework for the context questionnaires was published together with the assessment frameworks for mathematics, reading, science, problem solving and financial literacy. Therefore this section provides a summary of the context questionnaire framework only, with the interested reader referred to further details in OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. The framework for the context questionnaires in PISA 2012 outlines how PISA can be developed further as a sustainable database for educational policy and research. To this end, the framework starts with a review of the general purpose and policy relevance of PISA. Three types of policy-relevant “products” are identified:

- **Indicators** monitor the functioning, productivity and equity of education systems. PISA-based indicators refer to cognitive outcomes as well as non-cognitive outcomes such as attitudes, beliefs, motivation and learning-related behaviour, the latter being measured within the Student Questionnaire.
- PISA provides **knowledge** on individual, school and system-level factors that determine educational effectiveness. The programme reports representative, reliable data on factors that, according to previous research, are expected to impact student achievement. In addition to describing these factors, PISA estimates their direct and indirect relationships to student performance and other outcomes. Thus, it helps to understand how educational outcomes are produced.
- Each PISA assessment updates the sustainable, comparative **database** that allows researchers world-wide to study policy-oriented questions. PISA provides a data source for the study of educational contexts in general (e.g. how family, school and out-of-school education interact) and the study of educational variables in economic and sociological contexts (e.g. the relationship between demographics, economic wealth, economic growth and human resources).

Some of the relevant factors in understanding student performance, attitudes, and behaviours, and the functioning of education systems are straightforward (such as demographic variables, previous educational career choices, instructional time, and class size), some have been well established in previous PISA assessments (such as student socio-economic status, cognitive strategies, school-level decision-making), while others have proven to be less easily addressed within the PISA design (e.g. accountability policies at the system level, teacher variables, aspects of the classroom learning environment, or out-of-school activities). Choosing among the many variables that might be incorporated into the design is a complex process, directed by the priorities that countries have set for the study, but also informed by educational research.

In its Chapter 6 section on “The general knowledge base: Research in educational effectiveness”, the framework outline shows that the student questionnaire, the school questionnaire and the international options are rooted in well-established research instruments (OECD, 2013). Effectiveness factors can roughly be classified as being either input or



processes. Input factors are mostly related to the individual's social and personal background. Also, structural features like school size and funding are treated as inputs. Processes include learning and teaching as core processes with variables designed to capture their quantity and quality. Moreover, professional activities by teachers and principals as well as school policies and practices are classified as process variables. Figure 3.1 provides an overview of input, process, and outcome factors that are covered in the PISA 2012 Questionnaire Design.

■ Figure 3.1 ■

Taxonomy of educational outcomes and predictive factors

	Input	Processes	Outcomes
Students	Gender, grade level, socio-economic status	Attendance/truancy	Mathematical performance
	Educational career, grades	Outside-class activities - e.g. participation in after school programmes	Mathematics-related attitudes, beliefs and motivation
	Immigration background Family environment and support	Motivation, engagement	General school-related attitudes (towards learning outcomes and activities) and behaviour, e.g. commitment, truancy
	ICT ¹ experience, attitudes, skills	Learning and thinking strategies, test taking strategies	Learning motivation
	Openness, perseverance, problem solving styles	Learning time (including homework and private tuition)	
Classrooms	Class size, socio-economic background and ethnic composition	Opportunity to learn: Experience with various kinds of mathematical tasks, Concept familiarity Teaching practices: Teacher-directed instruction, student orientation, formative assessment and feedback Teaching quality: Classroom management/disciplinary climate, teacher support, cognitive activation	
	Teacher education/training, expertise	Instructional time, grouping practices	
Schools	Socio-economic background and ethnic composition Affluence of the community	Achievement orientation, shared norms, leadership, teacher morale and co-operation, professional development	Promotion/retention and graduation rates
	School funding, public vs. private School size	Admission and recruitment policies, tracking, course offerings/school curriculum, evaluation	Attendance
	Parental involvement	Teacher-student relations	
	Countries (Systems)	Economic wealth, social (in)equality	School funding, tracking and allocation, policies for professional teacher development, support for special needs and language minority students, hiring and certification policies
Diversity policies		Accountability and evaluation policies, locus of decision-making	

1. Information and Communication Technologies.

As PISA 2012 again has mathematics as its major domain, specific consideration has been given to issues of teaching and learning mathematics. This focus is present in three areas of the questionnaire design, as outlined in the Chapter 6 section of the Assessment Framework titled “*Learning conditions for mathematical literacy*” (OECD, 2013), namely non-cognitive outcomes, explanation of students’ intentions and behaviours related to mathematics and classroom teaching.

Non-cognitive outcomes: Measures of intrinsic and instrumental motivation for Mathematics, Learning Strategies (Control vs. Elaboration vs. Memorisation), self-efficacy, self-concept, and mathematics anxiety have been taken up from PISA 2003 after careful re-evaluation of their psychometric qualities.

Explaining student intentions and behaviour related to mathematics: How confident students are about their ability to solve mathematical tasks, as well as how students value mathematics, are highly relevant factors in predicting or explaining student behaviour with regard to mathematics, e.g. course-taking and career decisions. A number of expectancy value models both in psychology and in economics have been proposed to integrate both aspects of decision-making. One such model is Ajzen’s (1991) theory of planned behaviour, which states that volitional behaviour is determined by specific attitudes and subjective norms (= value component) plus perceived behavioural control (= expectancy component). In PISA 2012, a version of this model has been implemented in the Student Questionnaire. Students’ attitudes and attributions, perceptions of control, and subjective norms may predict their work ethics and intentions – e.g. their desire to spend time on mathematics homework – their study behaviour and finally their mathematics performance.



Classroom Teaching: PISA 2012 aims to identify country (and probably school) level profiles in opportunities to learn. Students were confronted with carefully crafted mathematics tasks – some representing mathematical abilities and content categories as mentioned in the PISA mathematics framework, some representing more traditional tasks asking for procedural and declarative knowledge. Following each of those items, students are asked to judge whether and how often they have seen similar tasks in their mathematics lessons and in previous assessments. These measures of content exposure are complemented by several scales describing teaching practices and teaching quality.

The framework's centrepiece, however, is its aim to map out a design for the PISA context questionnaires that will be sustainable well into the future (see Chapter 6 section "*Specifying the Questionnaire Design for PISA 2012*" in OECD, [2013]). To this end, the framework puts a system in place that accommodates recurring general material that is covered in every cycle and domain-specific material (for mathematics, science, or reading literacy, respectively), that is covered every third cycle, thus allowing for trend analyses of general as well as domain-specific issues. In addition, the framework's system also allows for thematic extensions and specific foci to enable PISA to anticipate and incorporate new material or topics of interest to its audience. The following types of measures are differentiated:

(I) General variables (for all cycles)

- Student-level inputs (grade, gender, parental education and occupation, family wealth, educational resources, cultural possessions, immigration status, heritage language, age on arrival in country, family support).
- School-level contexts and inputs (community size, resources, qualifications of teaching staff).
- School-level processes (decision-making, admission policies, assessment and evaluation policies, professional development, teacher engagement/morale, teacher-student relations, parental involvement).
- Instructional processes (learning time, disciplinary climate, teacher support).
- General non-cognitive outcomes – Commitment to learning (behavioural: truancy; personal goal: educational aspirations; motivational: learning engagement, affective: sense of belonging).

(II) Domain-specific trend variables (for major domain only, included every 9 years)

- Domain-specific non-cognitive outcome variables (strategies and metacognition, domain-related beliefs, self-related beliefs, motivation).
- Domain-specific processes variables (Opportunity To Learn, teaching practices, teaching quality, system- and school-level support).

(III) Thematic extension variables (extensions within individual cycles)

- International options (e.g. in 2012, educational career; ICT familiarity).
- Context variables for additional domains (e.g. ICT-related experiences relevant for computer-based problem solving).
- Descriptive and explanatory variables for specific reports (e.g. in 2012: mathematics-related motivations and intentions based on the theory of planned behavior).
- Malleable variables at the school level (e.g. tracking policies, teacher certification) that are specifically selected for descriptive purposes or for causal inference.

(IV) System-level data, mainly gathered outside of PISA

- Output of educational institutions (e.g. certificates).
- Financial and human resources invested into education.
- Access to and participation in education.
- Learning environment and organisation of schools.

An appropriate balance between (I), (II), (III), and (IV) is considered crucial for the overarching design of PISA questionnaires, and for the long term success of the PISA programme. In order to establish valid and reliable trends at the country level, it is important to implement a constant set of general variables in all cycles both for the calculation of proficiency estimates and as major reporting variables. Thus, these context and input background variables should not change. In order to provide trend information on non-cognitive outcomes and mathematics-related context/process variables, PISA 2012 retained as many variables that were used in the Student and School Questionnaires in 2003 as possible, unless they were shown not to work cross-culturally or not to account for differences in outcomes. Figure 3.2 provides an overview of the mathematics-specific indices in the student questionnaire that provided trend information between 2003 and 2012.



■ Figure 3.2 ■

Student Questionnaire – Mathematics-specific trend scales 2003-2012

Scale name 2003/2012	Scale label
General and mathematics processes	
BELONG*	Sense of belonging to school
STUREL	Student-teacher relations at school
DISCLIM	Disciplinary climate in the mathematics classroom
TEACHSUP	Teacher support in the mathematics classroom
Non-cognitive outcomes – Self and mathematics related cognitions	
ANXMAT	Mathematics anxiety
ATSCHL	Attitudes towards school: Learning outcomes
INSTMOT	Instrumental motivation to learn mathematics
INTMAT	Interest in and enjoyment of mathematics
MATHEFF	Mathematics self-efficacy
SCMAT	Mathematics self-concept

*This scale has been extended from 6 to 9 items in 2012. Trend analyses should only involve the 6 common items (i.e. ST87Q01 to ST87Q06).

CROSS-CULTURAL COMPARABILITY OF MEASURES IN THE CONTEXT QUESTIONNAIRES

One of the major challenges of an international study such as PISA is the cross-cultural validity and applicability of all instruments. In PISA 2012, the phenomenon that a number of non-cognitive student context constructs had been shown to be linked to performance in unexpected ways was given much thought and attention during the development phase of the context questionnaires (Kyllonen, Lietz and Roberts, 2010). More specifically, at the between-country level, data from previous cycles were such that countries with higher performance levels in a subject showed less positive attitudes towards that subject whereas more positive attitudes were recorded for lower-performing countries (Van de gaer and Adams, 2010; Van de gaer et al., 2012). Cross-cultural difference in response styles were considered to be – at least part of – the reason for this phenomenon.

Cross-cultural differences in response styles have been considered to represent a serious source of bias in international surveys that use Likert items. Several types of response styles – including extreme, central, acquiescent and disagreement response styles – have been described (e.g. Greenleaf, 1992; Clarke, 2000; Johnson et al., 2005). All of them can make it difficult to distinguish authentic cultural differences from “stylistic” biases in respondent behaviour (Van de Vijver and Poortinga, 1997; van Hemert, Poortinga and van de Vijver, 2007).

Proposed explanations of differences in response styles include the assumption of frame-of-reference effects whereby responses to attitude (or other) questions might differ systematically depending on which frame of reference (either across countries or across sub-groups within countries) is applied. These frames-of-reference include so-called “cultural macro values” (e.g. Hofstede, 2001; Schwartz, 2006; Triandis et al., 1988), the “Big Fish Little Pond Effect” (for an analysis using PISA 2000 data see Marsh and Hau, 2003), and social desirability (Holtgraves, 2004).

Three approaches, although intertwined, were identified in PISA 2012 to address this phenomenon. First, the phenomenon could be considered to reflect genuine differences between countries whereby some countries or cultural groups might have more positive attitudes regardless of the fact that the related actual context or outcome of interest is worse than in other countries. Second, it could be regarded as a measurement issue in that the measures or item types employed accentuate differences in response styles between countries and cultural groups. Therefore, it would be desirable to pursue measures that would be less affected by different response styles. Third, it could be considered that this phenomenon could be adjusted for through the application of different methods during the analysis stage (see, for example, Van de gaer and Adams, 2010).

In PISA 2012, the second approach was pursued further and four new item formats were introduced to the PISA 2012 Student Questionnaire, namely anchoring vignettes, signal detection debiasing based on the overclaiming technique, forced choice items, and Situational Judgment Tests (SJT).

Anchoring Vignettes

The first of the new methods was an alternative scoring of Likert-type items based on so-called anchoring vignettes (King and Wand, 2007; Hopkins and King, 2010). The anchoring vignettes approach has been used for making cross-country comparisons in various fields of research (Kapteyn, Smith and Van Soest, 2007; Salomon, Tandon and Murray, 2004; Kristensen and Johansson, 2008) but PISA 2012 was the first educational large-scale assessment to use the technique.



Two sets of so-called anchoring vignettes (see Figures 3.3 and 3.4) were included in the PISA 2012 Student Questionnaire to allow for alternative scoring of self-report items based on students' defined standards when using the 4-point agreement scale (strongly agree – agree – disagree – strongly disagree).

Each of these vignettes described behaviours of a hypothetical mathematics teacher that were indicative of lower or higher levels of classroom management (Figure 3.3) or teacher support (Figure 3.4), respectively. Each vignette combined several behavioural aspects. Students read the vignettes and were asked to indicate their level of agreement with a statement about the hypothetical teachers described in the vignettes. Differences in these ratings could be attributed to differences in the interpretation of the rating scale and general differences in preferred response behaviours as the underlying levels in the hypothetical teachers were held constant across countries.

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. The extension of the nonparametric scoring procedure (e.g., King and Wand, 2007) is described step by step in Chapter 17 of this report.

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. Results from analysis of Field Trial and Main Survey data showed that the vignettes capturing classroom management behaviours (see Chapter 17) produced clearer results (e.g. regarding the correct rank order of low, medium, and high vignettes by most respondents) and were better suited as anchors for students' self-report answers than the teacher support vignettes. In other words, a higher proportion of students did not give tied responses and the number of order violations – i.e., respondents' evaluations of the three anchors that violated the theoretically expected “correct” order – was lower for the classroom management vignettes than for the teacher support vignettes. These findings indicated that the former vignettes were worded in a way that made the difference between the high and low vignette larger than the latter vignettes.

■ Figure 3.3 ■

Anchoring vignettes based on classroom management behaviours

Low level	The students' in Mr. <name's> class frequently interrupt his lessons. As a result, he often arrives five minutes late to class.	ST84Q03
Medium level	The students' in Ms. <name's> class frequently interrupt her lessons. She always arrives five minutes early to class.	ST84Q01
High level	The students' in Ms. <name's> class are calm and orderly. She always arrives on time to class.	ST84Q02

Note. For each vignette students were asked to indicate how much they agree with the statement “Mr./Ms. <name> is in control of his/her classroom.”

■ Figure 3.4 ■

Anchoring vignettes based on teacher support behaviours

Low level	Ms. <name> sets mathematics homework once a week. She never gets the answers back to students before examinations.	ST82Q03
Medium level	Mr. <name> sets mathematics homework once a week. He always gets the answers back to students before examinations.	ST82Q02
High level	Ms. <name> sets mathematics homework every other day. She always gets the answers back to students before examinations.	ST82Q01

Note. For each vignette students were asked to indicate how much they agree with the statement “Mr./Ms. <name> is concerned about his students' learning.”

Topic Familiarity with Signal Detection Correction

The PISA 2012 student questionnaire includes several questions regarding familiarity with certain mathematics topics that were designed to measure students' opportunities to learn and content knowledge. When students are asked how well they know a given concept or whether they have seen a certain task type in their mathematics class, responses might, however, be affected by the same response tendencies that were revealed for other constructs.

One possible way of correcting for such response tendencies is the use of the so-called Overclaiming Technique (OCT; Paulhus, Harms, Bruce and Lysy, 2003; see also Zimmerman, Broder, Shaughnessy, and Underwood, 1977). This technique is a method that can be used to estimate both respondents' concept familiarity and their tendency to overstate what they know. It does this by collecting recognition judgments for intermixed concepts that actually exist, and foils, i.e. concepts that do not exist. In the PISA 2012 Student Questionnaire (ST62) this 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 13 actual mathematics concepts (e.g. “polynomial function”) and three foils (i.e. “proper number”, “subjunctive scaling” and “declarative fraction”). 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).

As discussed in Chapter 17, two indices were computed from students’ responses to this question (ST62). One index was a simple mean of students’ familiarity scores on the 5-point scale with the thirteen actual concepts (FAMCON). The other index took that mean and subtracted from it the mean familiarity score of the three foil concepts (FAMCONC).

Simple indices that can be derived are the so-called “Hit-Rate” and the “False-Alarm Rate”. From these, more complex indices of accuracy and bias could be derived based on Signal Detection Theory (SDT) approaches. Figure 3.5 gives an overview of all indices that were compared for the Field Trial. This figure also includes two additional indices, namely “topic familiarity” and “foil familiarity” that were calculated based on Field Trial data. These two additional indices are simple average scores derived from manifest student responses across all 16 items of the test.

■ Figure 3.5 ■

Overview of most prominent Signal Detection Measures and additional scoring rules for PISA 2012 questionnaire items

	Measure	Description/Formula
1	# hits	Number of real items rated as familiar
2	# misses	Number of real items rated as unfamiliar
3	# false alarms	Number of foils rated as familiar
4	# correct rejections	Number of foils rated as unfamiliar
5	H (hit rate)	Proportion of real items rated as familiar
6	F (false-alarm rate)	Proportion of foils rated as familiar
	P(c)	Percent correct (Hits + Correct Rejections)
7	z(H)	z-standardised hit rate
8	z(F)	z-standardised false-alarm rate
9	d' (“d prime”)	The number of hits relative to the number of false-alarms; $d' = z(H) - z(F)$
10	C (Bias)	$-0.5 * (z(H) - z(F))$
11	Topic Familiarity [†] (FAMCON)	Mean response for all concepts
12	Foil Familiarity [†]	Mean response for all foils
13	Adjusted Topic Familiarity [†] (FAMCONC)	Difference score: Topic Familiarity – Foil Familiarity

*Dependent on cut-off value

[†] alternative indices investigated in PISA Field Trial, not based on SDT

Situational Judgment Tests

Situational Judgment Test items (SJTs; Weekley and Ployhart, 2006) present short descriptions of situation with several possible responses which the test-taker must evaluate. There are many variations, but most often SJT items present several response options, and ask respondents to: (a) select the best option (multiple-choice); or, the best and the worst; (b) indicate for each option whether it would be acceptable or not (true-false), or (c) rate each option using a Likert scale. SJTs are widely used in industry and increasingly in education. In addition to the demonstrated validity of SJTs in employment settings (e.g. see McDaniel et al., 2001), SJTs have been shown as valid predictors in educational contexts such as performance during medical studies as well (e.g. Lievens et al., 2005). SJTs can reduce adverse impacts on, for example, mean score differences between racial groups as they tend to rely less on cognitive abilities than traditional item formats. Therefore, SJTs might be more appropriate instruments for minority groups than traditional tests.

Situational Judgment Tests were applied in the PISA 2012 Field Trial to measure two different constructs, namely Mathematics Motivation, and Problem Solving. Based on Field Trial results, only the Problem Solving SJT was retained for the Main Survey. The Motivation SJT did function reasonably well but could not add validity in terms of increasing hypothesised relationships with other relevant constructs beyond the traditional Likert scales.

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.



Forced Choice

If respondents have to decide between different alternatives, many of the problems associated with Likert scales can be avoided. In a so-called Forced-Choice assessment, a respondent is asked to choose one out of several alternative descriptions or assign ranks to all descriptions according to the extent to which the items describe the respondent's personality. Any ranking of n items can be coded equivalently using $[n(n-1)]/2$ binary outcome variables. For $n = 4$ choices {A, B, C, D}, the respondent has to assign ranking positions to each alternative, usually numbers from 1 (most preferred) to 4 (least preferred). The number of data-points that can be generated from this ranking is maximal but the cognitive load of such a comparison is also high. Alternatively, the respondent might be asked to indicate his or her most and least preferred option. This represents a partial ranking because it only assigns the first and the last ranks. The number of data points that can be generated is only somewhat smaller than for the full ranking, at the benefit of a reduction in cognitive load. A third alternative is to ask the test-taker to only choose his or her most (or least) preferred option. The simplest form of Forced Choice item is a paired comparison between only two choices. The ambiguity of the instruction is considerably reduced. Questions such as "Which of the two attributes describes you better?" or "Please rank the following 4 attributes according to how well they describe you" define much more clearly what the test taker has to do than the question "To what extent do you agree with each the following statements?" A drawback is that the cognitive load of the task increases when several response options have to be compared against each other.

Because it is impossible to endorse every item, the forced-choice format eliminates uniform biases such as acquiescence responding (Cheung and Chan, 2002), and can increase operational validity by reducing "halo" effects (Bartram, 2007). Forced Choice methods can reduce (but not fully eliminate) response biases. The reduction of bias is maximal when items in each block do not differ regarding their social desirability and other response styles, such as acquiescence or central tendency (e.g. Brown and Maydeu-Olivares, 2013).

Forced Choice (FC) assessments are more fake-resistant than Likert-type questionnaires. Especially when statements of equal social desirability are compared, faking becomes very complicated. Studies have shown reduced score inflation and maintenance of criterion related validities of FC measures in situations where examinees are motivated to fake (Bowen, Martin, and Hunt, 2002; Christiansen et al., 2005; White and Young, 1998).

Three important recent developments in psychometric models for FC data are the approaches by (a) Steve Stark, (b) Jimmy de la Torre, and (c) Anna Brown. Stark et al. (2005) proposed a model, the *Multi-Unidimensional Pairwise-Preference Model* (MUPP), for constructing and scoring multidimensional pairwise preference items. De la Torre et al. (2011) extended Stark et al.'s (2005) model by suggesting an item response model for preference data that can accommodate more than two components, and also different formats. They illustrate the application of the Bayesian Ipsative Data Analysis (BIDA) algorithm based on the MUPP model using Monte Carlo (MC) simulations. An alternative solution to the problem of ipsative data was presented by Brown and Maydeu-Olivares (2013) who suggest to transform ranking data into a series of paired comparisons. This transformed data can then be analysed with bi-factor models that account for the local independencies between the ranking based pairwise comparisons. The item characteristic function (ICF) for the binary outcome variable y , which is the result of comparing item i measuring trait a and item k measuring trait b , is then simply a standard two-dimensional normal ogive IRT (Item Response Theory) model for binary data with two exceptions: First, factor loadings are structured so that every binary outcome y_l involving the same item will share the same factor loading. Second, uniquenesses of latent response variables are structured so that they equal the sum of the 2 items involved. Third, the item characteristic functions are not independent, but patterned covariance matrices need to be specified.

A simpler approach is to derive a score for Forced Choice items based on the number of endorsements of one type of statements, i.e. an ipsative scoring strategy. This strategy is obviously inferior to the above described IRT models, but an alternative for small numbers of items, especially when the interest is not in deriving scores for all constructs but just the preference for one specific behaviour or attitude. This principle was used in the PISA 2012 Field Trial to measure students' preferences for Mathematics versus other subjects, as well as for their preferences for certain learning strategies.

TRANSITIONING THE CONTEXT QUESTIONNAIRES FROM PAPER ADMINISTRATION TO ONLINE ADMINISTRATION MODE

In addition to paper-based delivery of the questionnaires, PISA 2012 introduced an online administration mode for the School Questionnaire. On this first occasion, online administration of the School Questionnaire was optional for countries. Within countries, the possibility to print a pdf version of the School Questionnaire was provided, mainly to



enable school principals to obtain information that they had to get from elsewhere to answer some questions (e.g. about staff qualifications).

Nineteen countries and economies 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.

Improvements to the online School Questionnaire from the Field Trial to the Main Survey targeted international contractor processes and questionnaire functionalities for both National Centres and respondents, namely school principals or their deputies.

While the processes for the production of the online School Questionnaire were largely parallel to the processes for the production of paper-based questionnaires, a number of areas required additional work to support the transition to an online mode of administration:

- a) The creation of online source versions in English and French of the survey and survey architecture (e.g. online question construction; variable naming; validation rules; administration error messages). Online source versions were produced to:
 - help National Centres in the authoring of national versions of questionnaires; and
 - be incorporated in verification processes along with the negotiated Questionnaire Adaptation Spreadsheet (QAS) against the nationally adapted version of the questionnaire.
- b) The development of functionalities for the administration of the online School Questionnaire, which included:
 - import and management of country sampling frameworks in the online management interface; and
 - management of differentiated survey access between Consortium partners, National Centres and participating sampled schools.
- c) The development of new validity checks before and after implementation of the Main Survey. Online school questionnaire data were directly exported into *KeyQuest* – the data capture and cleaning software specifically developed for PISA – between consortium partners rather than data being exported from NCs into *KeyQuest* as with paper-based questionnaires. These checks included:
 - variable naming checks during the Final Check and linguistic verification processes before Main Survey implementation; and
 - validity reports that were run in *KeyQuest* of school sampling IDs from the online survey management interface.

Additional improvements for the Main Survey administration included: international contractor management of nationally adapted Field Trial questionnaires and Main Survey source updates in the online platform, improvements to the online authoring tool for National Centres, Consortium help and feedback during the authoring process.

UPDATE OF THE INTERNATIONAL STANDARD CLASSIFICATION OF OCCUPATIONS (ISCO) FROM ITS 1988 VERSION TO THE 2008 VERSION

Prior to PISA 2012 the 1988 version of the International Standard Classification of Occupations, ISCO-88, was used to code responses to open-ended questions by students about their mother's and father's occupation. In 2007, a new version, namely ISCO-08, was adopted by the International Labor Organisation (ILO) and recommended to be used by both the ILO and the European Commission (2009) in official statistics. The updated version covered more appropriately current occupations, particularly in the area of Information and Communication Technology and also defined more clearly different managerial levels. Hence, it was decided to adopt the ISCO-08 classification in PISA 2012.

In addition to including the nominal four-digit ISCO codes, the PISA dataset also include a mapping of ISCO onto an assumed interval scale – *International Socio-Economic Index of Occupational Status (ISEI)* which has been developed as a scale that is reflective of socio-educational status and is comparable across countries (Ganzeboom, 2010; Ganzeboom and Treiman, 2003). Together with information on parental education and home possessions, ISEI is subsequently used to create the *PISA index of economic, social and cultural status (ESCS)*. The rationale for using these three components is that socio-economic status is usually seen to be based on education, occupational status and income.



ESCS is used in many PISA reports and analyses, both as a control for the socio-economic status of students and schools and in bivariate correlations with performance as one of the main indicators of equity in an education system. Hence, the Consortium undertook analyses to examine the impact of the change from ISCO-88 to ISCO-08.

ESCS-88 is used as a label for the ESCS index that involves using ISEI values computed based on the ISCO-88 and ISEI-88. ESCS-08 is used as a label for the ESCS index that involves using ISEI values computed based on the ISCO-08 and ISEI-08.

To support the change from ESCS-88 to ESCS-08 a range of analyses were undertaken to document the implications of the update in terms of means, distributions of ESCS as well as the relationship between ESCS and student performance using Main Survey data from PISA 2012. Secondly, analyses aimed at exploring whether the changes in the ISCO classification have had implications for particular codes using data from the double coding process of the PISA 2012 Field Trial were undertaken. These results are reported in Chapter 17.

THE MEASUREMENT OF OPPORTUNITY TO LEARN

Current research on effective teaching (e.g., the comprehensive review of international teaching effectiveness research written by Good, Wiley and Florez for the 2009 edition of the *International Handbook of Research on Teachers and Teaching*) uses three kinds of measures to describe the classroom learning environment, namely measures of content, teaching practices and teaching quality.

Aspects of content matter, how it is selected, structured, and presented, have often been treated under the heading of Opportunity to Learn (OTL). The breadth and depth of content are described, coherence is rated and the alignment between intended curriculum (i.e. stated standards, syllabi) and implemented curriculum (i.e. the content actually taught) is evaluated. Schmidt and Maier (2009) argue that OTL is a rather straightforward concept: “What students learn in school is related to what is taught”, and they suggest to focus on OTL “in the narrowest sense: Student’s content exposure”.

Another set of measures refers to specific practices that are used by teachers, such as teacher-directed and student-directed activities, or various kinds of assessments. A well-known overview of evidence on teaching practices is provided by Hattie (2008). The OECD’s Teaching and Learning International Survey (TALIS) asked teachers about the frequency of using 13 different teaching practices which could be grouped into three dimensions: structuring practices (e.g. “I explicitly state learning goals.”), student-oriented practices (e.g. “Students work in small groups to come up with a joint solution to a problem or task.”), and enhanced activities (“Students work on projects that require at least one week to complete.”). These three dimensions could be identified across cultures (OECD, 2009).

Third, classroom environments have been characterised by aspects of the quality of teaching, i.e. how teachers deliver content and practices in the classroom. According to Pianta and Hamre (2009) who developed one of the most influential protocols for classroom observations, three dimensions are underlying the quality of teaching, namely classroom organisation, emotional support, and instructional support. This model has gained support from studies of teachers (Tschannen-Moran and Woolfolk-Hoy, 2007) as well as classroom research elsewhere (see Klieme, Pauli and Reusser, 2009; Baumert et al., 2010 who use the term “cognitive activation” rather than “instructional support”).

Opportunity to Learn, teaching practices, and quality measures may be combined to describe and evaluate classroom teaching and learning across cultures. All three kinds of measures have been implemented in the PISA 2012 Student Context Questionnaire to obtain information regarding the learning environment for mathematics.

Sometimes, the label “Opportunity to Learn” is used as embracing all aspects of instruction experienced by the student (e.g., Stevens, 1993). The PISA 2012 Questionnaire Framework, however, defines OTL as “coverage of content categories and problem types” to differentiate it from teaching practices and quality of teaching (OECD, 2013).

In PISA, the measurement of OTL has to be modified from approaches used in other studies, as the mathematics assessment is not framed according to content elements, but refers to fundamental mathematical abilities and broad content categories. Therefore, the measurement of OTL is based mainly on student judgements.

Opportunity to Learn content

Opportunity to Learn – in the sense of mathematical content that students experience – was assessed in PISA 2012 in three ways as detailed below.



Experience with mathematical tasks (ST61)

This question asked students how often they encounter various types of mathematical tasks during their time in school. Here, two subscales were formed from the same list of nine tasks that was used to measure of student self-efficacy in mathematics, a) experience with pure mathematical tasks (EXPUREM: ST61Q05, ST61Q07, ST61Q09), and b) experience with applied mathematical tasks (EXAPPLM: ST61Q01, ST61Q02, ST61Q03, ST61Q04, ST61Q06, ST61Q08).

Familiarity with mathematical concepts (ST62)

This question asked students to judge how familiar they were with 13 mathematical concepts. The response scale had five options: never heard of it (1), heard of it once or twice (2), a few times (3) or often (4), know it well and understand the concept (5).

All 13 items were combined into an overall index, Familiarity with Mathematics Concepts (FAMCON). Based on the 13 items, two indexes were used in reporting (see OECD 2014), a) Index of familiarity with algebra (ST62Q01, ST62Q03, and ST62Q05) and b) Index of Familiarity with Geometry (ST62Q06, ST62Q11, ST62Q13, and ST62Q15) but not included in the international database.

Question ST62 also included three foils, i.e. non-existing pseudo-concepts (ST62Q04, ST62Q10, and ST62Q12). If students indicated they heard of these or even know them well, this indicated overclaiming. The familiarity measure could be adjusted for the tendency to overclaim (see earlier explanation under “Topic Familiarity with Signal Detection Correction”); the adjusted index, Familiarity with Mathematics Concepts – Corrected for Overclaiming (FAMCONC) has been included in the international database.

Exposure to types of mathematical tasks in lessons and in tests (ST73-ST76)

Students were exposed to carefully crafted mathematics tasks – some representing applied mathematical reasoning as assessed in the PISA mathematics test, some representing inner-mathematical reasoning such as proofs and geometrical constructions, some representing short, well-defined word problems as frequently used in textbooks, or tasks checking procedural knowledge. For each of these four types of mathematical tasks, a short characterisation and two examples from different areas of mathematics are provided. Students were instructed not to solve these tasks. Instead, they were asked to recall how often they had previously encountered similar tasks in a) their mathematics lessons and b) in assessments on choosing one of four response options, namely “frequently”, “sometimes”, “rarely” or “never”.

As a result, two variables, one indicating the frequency in mathematics lesson and one indicating the frequency experienced in tests for the following four types of mathematical tasks:

- OTL - Algebraic Word Problem (ST73)
- OTL - Procedural Task (ST74)
- OTL - Pure Math Reasoning (ST75)
- OTL - Applied Math Reasoning (ST76)

Opportunity to Learn teaching practices

To operationalise this component of OTL, the teaching practices items from the OECD TALIS survey were adapted for use in PISA 2012. The items were reframed for use with students and some practices that are specific to mathematics were added. After some items were removed based on results in the Field Trial, 13 teaching practices remained in question ST79 which formed the following three scales:

- Teacher behaviour, Teacher-directed instruction (TCHBEHTD) based on items ST79Q01, ST79Q02, ST79Q06, ST79Q08, ST79Q15;
- Teacher behaviour, Student orientation (TCHBEHSO) based on items ST79Q03, ST79Q04, ST79Q07, ST79Q10; and
- Teacher behaviour, Formative assessment (TCHBEHFA) based on items ST79Q05, ST79Q11, ST79Q12, ST79Q17.

Opportunity to Learn teaching quality

As mentioned above, current research on teaching suggests that (a) classroom organisation and management, (b) teacher emotional and social support, and (c) cognitive activation have to be addressed as basic dimensions of instructional quality.



Two of these dimensions were covered in PISA 2003 and the respective scales continued to be used in 2012:

- Disciplinary climate (DISCLIMA), based on all five items in ST81, indicating *problems* with classroom organisation; and
- Mathematics teaching (TEACHSUP), based on all five items in ST77.

Response options for the items in both scales were “every lesson”, “most lessons”, “some lessons” and “never or hardly ever”.

The third dimension, cognitive activation (COGACT), based on nine items in ST80, is new to PISA 2012. Students were asked the extent to which they felt challenged by the tasks set by their mathematics teacher (e.g., “We usually have to think for a while in order to solve the problems we are assigned by our mathematics teacher”). This scale was used previously as a national option in PISA 2003 in Germany (see Baumert et al., 2008).

To test the usefulness of anchoring vignettes for adjusting non-cognitive scales for cross-cultural differences in response style in survey such as PISA, two scales measuring the quality of mathematics teaching were used.

The first scale, namely Teacher Support (MTSUP), consisted of one new item (ST83Q01 “My teacher lets us know we need to work hard”) plus three of the five items in ST77 that were used in the scale Mathematics teaching (TEACHSUP) but with changed response options, namely “strongly agree”, “agree”, “disagree” and “strongly disagree”. The corresponding anchoring vignette consisted of three items in ST82 that described three teachers in terms of the frequency of setting and returning homework.

The second scale, namely Classroom Management (CLSMAN), consisted of three items that were akin to the items in the Disciplinary climate (DISCLIMA) scale plus one item (ST85Q04) that was taken verbatim from that scale (i.e. ST81Q03). The corresponding anchoring vignette consisted of three items in ST84 that described three teachers of different levels of punctuality for lessons and student behaviour in class. Further details regarding the use of anchoring vignettes have been provided in an earlier section of this chapter.

THE ROTATION OF THE STUDENT CONTEXT QUESTIONNAIRE – DESIGN AND INTENDED ANALYSES

Whereas rotation of cognitive skills tests has been used extensively to increase content coverage of assessed domains for a long time, rotated student context questionnaires were used for the first time in a main data collection of an international comparative assessment in education in PISA 2012. This was done to increase the content coverage of topics of interest to PISA in the questionnaire without increasing the response time for individual students to more than 30 minutes.

The rotated design was such that three forms of the questionnaire contained a common part and a rotated part. The common part, which was administered to all students, contained questions to obtain information about gender, language at home, migrant background, home possessions, parental occupation and education. The rotated part which was administered to one-third of students contained questions about attitudinal and other non-cognitive constructs.

Prior to going down the path of using rotated student questionnaires in the main data collection, extensive analyses were undertaken to examine the impact of this methodology on the continuity of the results. Thus, PISA 2006 data for nine heterogeneous countries were rescaled after having been restructured to simulate the outcomes of the use of different rotated context questionnaire designs. Results revealed negligible differences when means, standard deviations, percentiles were estimated using plausible values drawn with multilevel item response models that adopted different approaches to questionnaire rotation. Also, only 110 of 2 700 correlations between student context constructs and proficiency differed by more than 0.03 with standard errors increasing either not at all or by 0.01 (Adams, Lietz and Berezner, 2013).

The logistics of questionnaire administration became slightly more complex by using a rotated Student Questionnaire design for several reasons. First, more adjustments needed to be negotiated between National Project Managers (NPMs) and the Consortium. Second, although the absolute number of student questionnaires to be printed remained the same for a given sample size, different forms had to be printed, increasing production costs. Third, during administration, about the same number of students had to respond to randomly assigned Student Questionnaire forms which remained



relatively simple as Student Questionnaire forms were not linked to specific cognitive test forms. Despite these slightly more complex logistics, the rotated Student Questionnaire was administered successfully in the great majority of participating countries. This was not least due to the experience with the administration of rotated forms of the Student Questionnaire in all PISA Field Trials to date.

The finally chosen design as illustrated in Figure 3.6 was a rotation with constructs being asked in two of the three forms to allow joint analyses of these constructs. This resulted in responses from two third of students per construct but freed up less space. Still, it was considered preferable as a full covariance matrix could be derived as every construct was asked with every other construct at least once.

■ Figure 3.6 ■

Final design of rotated Student Context Questionnaires in PISA 2012

Form A	Form B	Form C
Common part (8 minutes)		
Rotated question set 1 (11 minutes)	Rotated question set 3 (11 minutes)	Rotated question set 3 missing
Rotated question set 2 (11 minutes)	Rotated question set 3 missing	Rotated question set 2 (11 minutes)
Rotated question set 3 missing	Rotated question set 1 (11 minutes)	Rotated question set 3 (11 minutes)

Notes: Three rotated forms, two-thirds of students answer questions in rotated parts.

As can be seen in Figure 3.6, the rotated student context questionnaire in PISA 2012 consisted of two parts, namely the “common” and the “rotated” part. Questions in the “common” part were answered by all students while questions in the “rotated” parts were answered by two thirds of the student sample.

It should be noted that each rotated question set occurred first in one of the forms in order to balance the possibility of missing data due to respondents’ fatigue in the latter part of the questionnaire. Thus, as can be seen in Figure 3.6, Form A contained question set 1 first, question set 3 features first in Form B while in Form C students were first asked to respond to question set 2.

The common part was estimated to take students about eight minutes to complete. Each rotated question set could take up about eleven minutes of response time. The common part and two rotated question sets, then, resulted in the usual 30 minute response time to the Student Questionnaire for an individual student. Timing estimates were derived from knowledge gained from previous PISA cycles as well as cognitive laboratories during the item development and Field Trial phase.

The content in the common part (see Figure 3.7) included demographics questions and major reporting variables.

For the rotated parts of the Student Questionnaire, the following guiding principles were applied in the allocation of questions to the three question sets:

- Use intact scales only. Do not split items constituting a construct across forms.
- Allocate questions with similar themes to a question set.
- Each question set not to exceed 11 minutes; question sets should be of similar length.
- Balance constructs in terms of their correlation with performance. In other words, on average, correlation with performance of constructs in question sets should be similar based on results of Field Trial.

Three question sets were designed in this way whereby question set 1 was included in Forms A and B. Question set 2 was included in Forms A and C. Question set 3 was included in Forms B and C (see Figure 3.7). Details regarding the questions in the rotated part of the three Student Questionnaire forms are given in Figure 3.9.

Question set 1 contained items covering attitudes towards mathematics and the problem solving Situational Judgement Test items. Question set 2 included items on school climate and attitudes towards school. Mathematics anxiety was also included in question set 2 although, conceptually, it would have been place more appropriately in question set 1. However, as items in question set 1 already showed reasonable correlations with performance while correlations between items and performance were a bit weaker in question set 2, mathematics anxiety was placed in question set 2 due to its relatively higher correlation with mathematics performance. Question set 3 consisted of items measuring Opportunity to Learn and learning strategies. Overall, question 3 was slightly shorter than question sets 1 and 2 but only marginally so.

■ Figure 3.7 ■
Questions in the common part

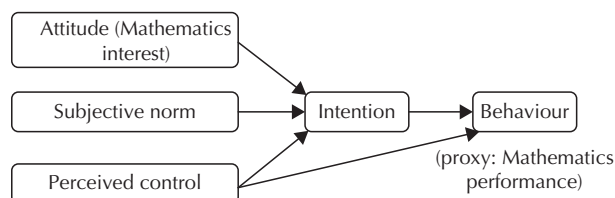
Question number	Description
ST01	Grade
ST02	Country study programme
ST03	Age of student
ST04	Sex of student
ST05	Attend <ISCED 0>
ST06	Age at <ISCED 1>
ST07	Grade Repeating
ST08	Truancy; Times late for school
ST09	Truancy; Days unexcused absence
ST115	Truancy; Times skipped classes
ST11	Family structure
ST12	Mother's occupation (ISCO); Component of ESCS
ST13	Mother's educational level – Schooling (ISCED); Component of ESCS
ST14	Mother's educational level – Post school (ISCED); Component of ESCS
ST15	Mother's current job status; Component of ESCS
ST16	Father's occupation (ISCO); Component of ESCS
ST17	Father's educational level (ISCED) – Schooling (ISCED); Component of ESCS
ST18	Father's educational level (ISCED) – Post school (ISCED); Component of ESCS
ST19	Father's current job status; Component of ESCS
ST20	Immigrant background
ST21	Age of arrival in test country
ST25	Language spoken at home
ST26	General home possessions plus country-specific wealth items; Component of ESCS
ST27	Number of certain possessions in household; Component of ESCS
ST28	Books at home

Notes: Questions are listed in the order in which they appeared in the Student Questionnaire. Question numbers in the Field Trial and Main Survey were the same which meant that some numbers are missing (e.g. ST22 to ST24) because they were deleted after the trial from the Student Questionnaire (some countries included these questions in the Educational Career Questionnaire). Also, some numbers were new (e.g. ST115) as some questions were revised substantially after the Field Trial.

Analysts interested in exploring approaches to problem solving in the Student Questionnaire – maybe in order to relate these approaches to proficiency in problem solving from the cognitive tests – are pointed to question set 1 which covers this area (ST94, ST96, ST101 and ST104).

Some scales in the PISA 2012 context questionnaire framework and subsequent questionnaire were designed to enable the exploration of the Theory of Planned Behaviour (Ajzen, 1991; Armitage and Conner, 2001; Sheeran, 2002). Analysts interested in exploring a model derived from this theory such as the one presented below, should also turn to question set 1.

■ Figure 3.8 ■
Model of Theory of Planned Behaviour





■ Figure 3.9 ■
Questions in the rotated parts

FORM A		FORM B		FORM C	
Q. number	Description	Q. number	Description	Q. number	Description
ST01-28	Common part (see Figure 3.7)	ST01-28	Common part (see Figure 3.7)	ST01-28	Common part (see Figure 3.7)
ST29	Instrumental Motivation (Q2, 5, 7, 8) Mathematics Interest (Q1, 3, 4, 6)	ST42	Mathematics Self-Concept (Q2, 4, 6, 7, 9); Mathematics Anxiety (Q1, 3, 5, 8, 10)	ST53	Learning Strategies (Self-Control)
ST35	Subjective Norms	ST77	Teacher Support in Mathematics Class	ST55	Out-of-School- Lessons
ST37	Mathematics Self-Efficacy	ST79	Teaching Practices	ST57	School Study Time
ST43	Perceived Control of Mathematics Performance	ST80	Cognitive Activation in Mathematics Lessons	ST61	Experience with Applied Maths Tasks (Q1-4, 6, 8); Experience with Pure Math tasks (Q5, 7, 9)
ST44	Attributions to Failure in Mathematics	ST81	Disciplinary Climate	ST62	Familiarity with Maths Concepts
ST46	Mathematics Work Ethic	ST82	Anchoring Vignettes - Teacher Support	ST69	Minutes in <Class Period>
ST48	Mathematics Intentions (Forced-Choice)	ST83	Mathematics Teacher Support	ST70	Number of <Class Period> per Week
ST49	Mathematics Behaviour	ST84	Anchoring Vignettes - Classroom Management	ST71	Number of All <Class Period> per Week
ST93	Perseverance	ST85	Mathematics Teacher's Classroom Management	ST72	Class Size
ST94	Openness for Problem Solving	ST86	Student-Teacher Relations	ST73	OTL - Algebraic Word Problem
ST96	Problem Solving Strategies (SJT-Text Message)	ST87	Sense of Belonging to School	ST74	OTL - Procedural Task
ST101	Problem Solving Strategies (SJT-Route Selection)	ST88	Attitude towards School: Learning Outcomes	ST75	OTL - Pure Mathematics Reasoning
ST104	Problem Solving Strategies (SJT-Ticket Machine)	ST89	Attitude towards School: Learning Activities	ST76	OTL - Applied Mathematics Reasoning
ST53	Learning Strategies (Self-Control)	ST91	Perceived Control of Success in School	ST42	Mathematics Self-Concept (Q2,4,6,7,9); Mathematics Anxiety (Q1,3,5,8,10)
ST55	Out-of-School-Lessons	ST29	Instrumental Motivation (Q2, 5, 7, 8) Mathematics Interest (Q1, 3, 4, 6)	ST77	Teacher Support in Mathematics Class
ST57	School Study Time	ST35	Subjective Norms	ST79	Teaching Practices
ST61	Experience with Applied Mathematics Tasks (Q1-4, 6, 8); Experience with Pure Mathematics tasks (Q5, 7, 9)	ST37	Mathematics Self-Efficacy	ST80	Cognitive Activation in Mathematics Lessons
ST62	Familiarity with Mathematics Concepts	ST43	Perceived Control of Mathematics Performance	ST81	Disciplinary Climate
ST69	Minutes in <Class Period>	ST44	Attributions to Failure in Mathematics	ST82	Anchoring Vignettes - Teacher Support
ST70	Number of <Class Period> per Week	ST46	Mathematics Work Ethic	ST83	Mathematics Teacher Support
ST71	Number of All <Class Period> per Week	ST48	Mathematics Intentions (Forced-Choice)	ST84	Anchoring Vignettes - Classroom Management
ST72	Class Size	ST49	Mathematics Behaviour	ST85	Mathematics Teacher's Classroom Management
ST73	OTL - Algebraic Word Problem	ST93	Perseverance	ST86	Student-Teacher Relations
ST74	OTL - Procedural Task	ST94	Openness for Problem Solving	ST87	Sense of Belonging to School
ST75	OTL - Pure Mathematics Reasoning	ST96	Problem Solving Strategies (SJT-Text Message)	ST88	Attitude towards School: Learning Outcomes
ST76	OTL - Applied Mathematics Reasoning	ST101	Problem Solving Strategies (SJT-Route Selection)	ST89	Attitude towards School: Learning Activities
		ST104	Problem Solving Strategies (SJT-Ticket Machine)	ST91	Perceived Control of Success in School



Note

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.

References

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).

Ajzen, I. (1991), “The theory of planned behaviour”, *Organizational Behavior and Human Decision Processes*, No. 50, pp. 179-211.

Armitage, C.J. and M. Conner (2001), “Efficacy of the theory of planned behaviour: A meta-analytic review”, *British Journal of Social Psychology*, No. 40, pp. 471-499.

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>

Baumert, J., M. Kunter, W. Blum, M. Brunner, T. Voss, A. Jordan, U. Klusmann, S. Krauss, M. Neubrand and Y.-M. Tsai (2010), “Teachers’ Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress”, *American Educational Research Journal*, Vol. 47, pp. 133-180.

Baumert, J., W. Blum, M. Brunner, T. Dubberke, A. Jordan, U. Klusmann, S. Krauss, M. Kunter, K. Löwen, N. Neubrand and Y.-M. Tsai (2008), “Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV) - Dokumentation der Erhebungsinstrumente (Baumert et al. [2009])”, Materialien aus der Bildungsforschung, No. 83, Max-Planck-Institut für Bildungsforschung, Berlin.

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. http://dx.doi.org/10.1207/S15328007SEM0901_4

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

Clarke, I., III. (2000), “Extreme response style in cross-cultural research: An empirical investigation”, *Journal of Social Behavior and Personality*, No. 15, pp. 137-152.

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, Illinois.

Ganzeboom, H.B.G. (2010), “Do’s and Don’ts of Occupation Coding with an Extension on ISCO-08”, Working Paper of the Department of Social Research Methodology, VU Amsterdam, last revised January 3, 2010.

Ganzeboom, H. B.G. and D. J. Treiman (2003), “Three Internationally Standardised Measures for Comparative Research on Occupational Status”, in Hoffmeyer-Zlotnik, J. H.P. and C.Wolf (eds.), *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*, Kluwer Academic Press New York, pp. 159-193.

Good, T. L., C. R. H. Wiley and I. R. Florez (2009), “Effective teaching: an emerging synthesis”, in G. Dworkin (ed.), *International Handbook of Research on Teachers and Teaching*, Springer, New York, pp. 803–816.

Greenleaf, E.A. (1992), “Improving rating scale measures by detecting and correcting bias components in some response styles”, *Journal of Marketing Research*, No. 24, pp. 176-188.

Hattie, J. (2008), *Visible learning: A synthesis of over 800 Meta-analyses Relating to Achievement*, Routledge.



Hofstede, G. (2001), *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*, Sage, Thousand Oaks, California.

Holtgraves, T. (2004), "Social desirability and self-reports: Testing models of socially desirable responding", *Personality and Social Psychology Bulletin*, No. 30, pp. 161-172.

Hopkins, D.J. and G. King (2010), "Improving Anchoring Vignettes: Designing Surveys to Correct for Interpersonal Incomparability", *Public Opinion Quarterly*, Vol. 74, pp. 201-22.

International Labour Organisation (2008), *International Standard Classification of Occupations: ISCO-08*, International Labour Organisation, Geneva.

Johnson, T., P. Kulesa, Y. I. Cho and S. Shavitt (2005), "The relation between culture and response styles: Evidence from 19 countries", *Journal of Cross-Cultural Psychology*, No. 36, 2, pp. 264-277.

Kapteyn, A., J.P. Smith, A. Van Soest and H. Vonkova (2007), "Vignettes and Self-Reports of Work Disability in the US and the Netherlands", *American Economic Review*, Vol. 97, pp. 461-73.

King, G. and J. Wand (2007), "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes", *Political Analysis*, Vol. 15, No. 1, pp. 46-66.

Klieme, E., C. Pauli and K. Reusser (2009), "The Pythagoras Study: Investigating Effects of Teaching and Learning in Swiss and German Classrooms", in Janik T. and T. Seidel (eds.), *The Power of Video Studies in Investigating Teaching and Learning in the Classroom*, Waxmann Verlag, Münster, pp. 137-160.

Kristensen, N. and E. Johansson (2008), "New Evidence on Cross-Country Differences in Job Satisfaction Using Anchoring Vignettes", *Labour Economics*, Vol. 15, pp. 96-117.

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

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

Marsh, H.W. and K.T. Hau (2003), "Big-Fish-Little-Pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools", *American Psychologist*, Vol. 58(5), May 2003, pp. 364-376.
<http://dx.doi.org/10.1037/0003-066X.58.5.364>

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.

OECD (2014), *PISA 2012 Results: What Students Know and Can Do - Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014)*, PISA, OECD Publishing, Paris.
<http://dx.doi.org/10.1787/9789264208780-en>

OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD Publishing, Paris, pp. 167-200.
<http://dx.doi.org/10.1787/9789264190511-en>

OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris.
<http://dx.doi.org/10.1787/9789264068780-en>

Pianta, R.C. and B.K. Hamre (2009), "Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity", *Educational Researcher*, No. 38, pp. 109-119.
<http://dx.doi.org/10.3102/0013189X09332374>

Paulhus, D. L., P. D. Harms, N. M. 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>

Salomon, J. A., A. Tandon and C.J.L. Murray (2004), "Comparability of Self-Rated Health: Cross-Sectional Multi-Country Survey Using Anchoring Vignettes", *British Medical Journal*, Vol. 328, pp. 258-61.

Schwartz, S.H. (2006), "Value orientations: Measurement, antecedents and consequences across nations", in Jowell, R., C. Roberts, R. Fitzgerald and G. Eva (eds.), *Measuring Attitudes Cross-nationally - Lessons from the European Social Survey*, Sage, London.

Sheeran, P. (2002), "Intention-behaviour relations: a conceptual and empirical review", in Strobe W. and M. Hewstone (eds.), *European Review of Social Psychology*, Vol. 12, Wiley, Chichester, pp. 1-30.

Schmidt, W.H. and A. Maier (2009), "Opportunity to Learn", in Sykes G., B. Schneider and D.N. Plank (eds.), *Handbook of Education Policy Research*, Routledge, New York, pp. 541-559.

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.

Stevens, F. (1993), "Applying an Opportunity-to-Learn Conceptual Framework to the Investigation of the Effects of Teaching Practices via Secondary Analyses of Multiple-Case-Study Summary Data", *Journal of Negro Education*, Vol. 62, (3), pp. 232-248.

Tschannen-Moran, M. and A. Woolfolk-Hoy (2007), "The differential antecedents of self-efficacy beliefs of novice and experienced teachers", *Teaching and Teacher Education*, Vol. 23, pp. 944-956.

Triandis, H.C., R.Bontempo, M.J.Villareal, M. Asai and N. Lucca (1988), "Individualism and Collectivism: Cross-Cultural Perspectives on Self - Ingroup Relationships", *Journal of Personality and Social Psychology*, No. 54, (2), pp. 323-338.

Van de gaer, E. and R.J. Adams (2010), "The Modeling of Response Style Bias: An Answer to the Attitude-Achievement Paradox?", Paper presented at the annual meeting of the American Education Research Association (AERA), 30 April - 4 May, 2010, Denver, Colorado.

Van de gaer, E., A.Grisay, W.Schulz and E. Gebhardt (2012), "The reference group effect: An explanation of the paradoxical relationship between academic achievement and self-confidence across countries", *Journal of Cross-Cultural Psychology*, No. 43(8), pp. 1205-1228

Van de Vijver, F.J.R. and Y.H. Poortinga, (1997), "Towards an integrated analysis of bias in cross-cultural assessment", *European Journal of Psychological Assessment*, Vol. 13, (1), Hogrefe and Huber Publishers.

Van Hemert, D., Y.H. Poortinga and F.J.R. Van de Vijver (2007), "Emotion and culture: A meta-analysis", *Cognition and Emotion*, No. 21, pp. 913-943.

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, California.

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.