

## Chapter 2

# Using Mathematical Competencies to Predict Item Difficulty in PISA: A MEG Study

Ross Turner, John Dossey, Werner Blum, and Mogens Niss

**Abstract** This paper reports an analysis of features of mathematics assessment items developed for the OECD's Programme for International Student Assessment survey (PISA) in relation to a set of six mathematical competencies. These competencies have underpinned the PISA mathematics framework since the inception of the PISA survey; they have been used to drive mathematics curriculum and assessment review and reform in several countries; and the results of the study are therefore likely to be of interest to the broad mathematics education community.

We present a scheme used to describe this set of mathematical competencies, to quantify the extent to which solution of each assessment item calls for the activation of those competencies, and to investigate how the demand for activation of those competencies relates to the difficulty of the items. We find that the scheme can be used effectively, and that ratings of items according to their demand for activation of the competencies are highly predictive of the difficulty of the items.

**Keywords** Mathematical competencies • PISA • Assessment • Item characteristics

---

R. Turner, M.Sc., DipEdPsych (✉)  
Australian Council for Educational Research (ACER), 19 Prospect Hill Road,  
Private Bag 55, Camberwell, VIC 3124, Australia  
e-mail: turner@acer.edu.au

J. Dossey, Ph.D.  
Department of Mathematics (Emeritus), Illinois State University,  
65111 East Crystal Ridge, Tucson, AZ 85739, USA  
e-mail: jdossey@ilstu.edu

W. Blum  
Mathematics Education, Institute for Mathematics, University of Kassel,  
D-34109 Kassel, Germany  
e-mail: blum@mathematik.uni-kassel.de

M. Niss  
Mathematics and Mathematics Education, IMFUFA/NSM, Roskilde University,  
P.O. Box 260, DK 4000 Roskilde, Denmark  
e-mail: mn@ruc.dk

## 2.1 Introduction

What are the factors that influence the difficulty of PISA mathematics survey items? The publication of data from the PISA 2003 survey (OECD, 2004), when mathematics was the major survey domain, has enabled a deep study of cognitive factors that influence the difficulty of mathematics items. The framework on which that survey was based (OECD, 2003) outlines a set of mathematical competencies originally described in the work of Mogens Niss and his Danish colleagues (see Niss, 2003; Niss & Hoejgaard, 2011). Such an understanding of item difficulty has the potential to guide the construction of new items to better assess the full range of the PISA mathematics scale, as well as to enhance the reporting of student performance associated with PISA assessments.

To the extent that these “Niss competencies” have resonance in various national curricula (e.g. in Denmark; see the official guidelines from the Ministry of Education: [www.ug.dk/uddannelser/professionsbacheloruddannelse/enkeltfag](http://www.ug.dk/uddannelser/professionsbacheloruddannelse/enkeltfag)), have been used to evaluate curriculum outcomes and even have acted as drivers of curriculum and assessment reform (e.g. in Germany; see Blum, Druke-Noe, Hartung, & Köller, 2006, and in Catalonia, Spain; see Planas, 2010), an understanding of their influence on the difficulty of mathematics items will have far wider relevance than just within the PISA context, and will contribute more generally to an important area of knowledge in mathematics education.

The authors led an investigation that has extended over several years, beginning in October 2003. They built on earlier work aiming at understanding student achievement in mathematics developed by de Lange (1987), Niss (1999), and Neubrand et al. (2001). The investigation has focused on six mathematical competencies which are a re-configuration of the set of competencies which have been at the heart of the Mathematics Framework for PISA from the beginning (see OECD, 2003, 2006). These competencies describe the essential activities when solving mathematical problems and are regarded as necessary prerequisites for students to successfully engage in “making sense” of situations where mathematics might add to understanding and solutions. These six competencies were:

- Reasoning and argumentation
- Communication
- Modelling
- Representation
- Solving problems mathematically (referred to as Problem solving)
- Using symbolic, formal and technical language and operations (referred to as Symbols and formalism).

These competencies are not meant to be sharply disjoint. Rather, they overlap to a certain degree, and mostly they have to be activated jointly in the process of solving mathematical problems.

The initial investigation has consisted in developing operational definitions of these six competencies, and in describing four levels of demand for activation of each competency (see Sect. 2.2). PISA survey items have been analysed in relation

to those definitions and level descriptions, by the application of a set of rating values to each item for each competency. The resulting ratings have then been analysed as predictor variables in a regression on the empirical difficulty of the items, derived from the PISA 2003 survey data. The item ratings have been found to be highly predictive of the difficulty of the items (see Sect. 2.3). In addition, statistical studies have been conducted examining other variables, such as the four PISA mathematical content strands (quantity, space and shape, change and relationships, and uncertainty), the PISA contexts in which the item items are presented to students (personal, education/occupational, public, scientific, and intra-mathematical), as well as the item formats themselves (various forms and combinations of multiple-choice, closed constructed-response, and open constructed-response items). None of these studies showed that these variables, acting singly or in combination with one another, explained significant proportions of the variation observed in item difficulty.

In this paper, we will present those competency definitions and level descriptions as well as the essential outcomes of the analysis conducted.

## 2.2 The Competency Related Variables

The material following in Table 2.1 contains the definitions and difficulty level descriptions of the six mathematical competencies used in this investigation so far. Each of the six competencies has an operational definition bounding what constitutes the competency as it might appear in PISA mathematics assessment items and then four described levels (labeled as levels 0, 1, 2, and 3) of each variable.

## 2.3 Analysis of the Application of the MEG Item Difficulty Framework

The following analyses provide an examination of the efficacy of the MEG Item Difficulty Framework in explaining the variability present in student performance on the 48 items common to the PISA 2003 and PISA 2006 mathematics assessments. We examine this efficacy from a number of perspectives: correlation of variable code average values, coder consistency, percentage of variance explained, consistency across assessments, and factor structure.

### 2.3.1 Psychometric Quality

#### 2.3.1.1 Correlation of Variable Average Code Values

Table 2.2 contains the results of a correlation of the coding data associated with each of the six competency-based variables. Note that in this and subsequent tables, the competency labels are abbreviated as follows: REA for Reasoning and

**Table 2.1** MEG item-difficulty coding framework

Symbols and formalism	
Variable-definition	<b>Symbols and formalism</b> [Understanding, <b>manipulating</b> , and <b>making use</b> of symbolic expressions within a mathematical context (including arithmetic expressions and operations), governed by mathematical <b>conventions and rules</b> ; understanding and <b>utilising constructs</b> based on definitions, rules and <b>formal systems</b> .]
Level 0	No mathematical rules or symbolic expressions need to be activated beyond fundamental arithmetic calculations, operating with small or easily tractable numbers.
Level 1	Make direct use of a simple functional relationship, either implicit or explicit (for example, familiar linear relationships); use formal mathematical symbols (for example, by direct substitution or sustained arithmetic calculations involving fractions and decimals) or activate and directly use a formal mathematical definition, convention or symbolic concept.
Level 2	Explicit use and manipulation of symbols (for example, by algebraically rearranging a formula); activate and use mathematical rules, definitions, conventions, procedures or formulae using a combination of multiple relationships or symbolic concepts.
Level 3	Multi-step application of formal mathematical procedures; working flexibly with functional or involved algebraic relationships; using both mathematical technique and knowledge to produce results.
Reasoning and Argumentation	
Variable-definition	<b>Reasoning and argumentation</b> [Logically rooted thought processes that explore and link problem elements so as to <b>make inferences</b> from them, or to <b>check a justification that is given</b> or <b>provide a justification</b> of statements.]
Level 0	Make direct inferences from the instructions given.
Level 1	Reflect to join information to make inferences, (for example to link separate components present in the problem, or to use direct reasoning within one aspect of the problem).
Level 2	Analyse information (for example to connect several variables) to follow or create a multi-step argument; reason from linked information sources.
Level 3	Synthesise and evaluate, use or create chains of reasoning to justify inferences or to make generalisations, drawing on and combining multiple elements of information in a sustained and directed way.
Problem solving	
Variable-definition	<b>Solving problems mathematically</b> [Selecting or devising, as well as implementing, a mathematical strategy to solve problems arising from the task or context.]
Level 0	Take direct actions, where the strategy needed is stated or obvious.
Level 1	Decide on a suitable strategy that uses the relevant given information to reach a conclusion.
Level 2	Construct a strategy to transform given information to reach a conclusion.
Level 3	Construct an elaborated strategy to find an exhaustive solution or a generalised conclusion; evaluate or compare strategies.

(continued)

**Table 2.1** (continued)

Modelling	
Variable-definition	<b>Modelling</b> [ <b>Mathematising</b> an extra-mathematical situation (which includes structuring, idealising, making assumptions, building a model), or <b>making use</b> of a given or constructed model by <b>interpreting</b> or validating it in relation to the context.]
Level 0	Either the situation is purely intra-mathematical, or the relationship between the real situation and the model is not needed in solving the problem.
Level 1	Interpret and infer directly from a given model; translate directly from a situation into mathematics (for example, structure and conceptualise the situation in a relevant way, identify and select relevant variables, collect relevant measurements, make diagrams).
Level 2	Modify or use a given model to satisfy changed conditions or interpret inferred relationships; or choose a familiar model within limited and clearly articulated constraints; or create a model where the required variables, relationships and constraints are explicit and clear.
Level 3	Create a model in a situation where the assumptions, variables, relationships and constraints are to be identified or defined, and check that the model satisfies the requirements of the task; evaluate or compare models.
Communication	
Variable-definition	<b>Communication</b> [Decoding and <b>interpreting</b> statements, questions and tasks; including <b>imagining</b> the situation presented so as to <b>make sense</b> of the information provided; <b>presenting and explaining</b> one's work or reasoning.]
Level 0	Understand a short sentence or phrase relating to a single familiar concept that gives immediate access to the context, where it is clear what information is relevant, and where the order of information matches the required steps of thought.
Level 1	Identify and extract relevant information. Use links or connections within the text that are needed to understand the context and task, or cycle within the text or between the text and other related representation/s. Any constructive communication required is simple, but beyond the presentation of a single numeric result.
Level 2	Use repeated cycling to understand instructions and decode the elements of the context or task; interpret conditional statements or instructions containing diverse elements; or actively communicate a constructed description or explanation.
Level 3	Create an economical, clear, coherent and complete description or explanation of a solution, process or argument; interpret complex logical relations involving multiple ideas and connections.
Representation	
Variable-definition	<b>Representation</b> [ <b>Interpreting</b> , translating between, and <b>making use</b> of given representations; <b>selecting</b> or <b>devising</b> representations to capture the situation or to present one's work. The representations referred to are depictions of mathematical objects or relationships, which include equations, formulae, graphs, tables, diagrams, pictures, textual descriptions, concrete materials]

(continued)

**Table 2.1** (continued)

Level 0	Directly handle a given representation, for example going directly from text to numbers, reading a value directly from a graph or table, where minimal interpretation is required in relation to the situation.
Level 1	Select and interpret one standard or familiar representation in relation to a situation.
Level 2	Translate between or use two or more different representations in relation to a situation, including modifying a representation; or devise a simple representation of a situation.
Level 3	Understand and use a non-standard representation that requires substantial decoding and interpretation; or devise a representation that captures the key aspects of a complex situation; or compare or evaluate representations.

**Table 2.2** Correlations of competency-based variable values based on the coding of 48 PISA mathematics items by eight coders

	AVGSYM	AVGREA	AVGPS	AVGMOD	AVGCOM
<b>AVGREA</b>	0.283 0.051				
<b>AVGPS</b>	0.301* 0.038	0.721* 0.000			
<b>AVGMOD</b>	0.606* 0.000	0.455* 0.001	0.401* 0.005		
<b>AVGCOM</b>	0.405* 0.004	0.471* 0.001	0.100 0.497	0.267 0.066	
<b>AVGREP</b>	0.062 0.676	0.314* 0.030	0.303* 0.036	0.261 0.073	0.082 0.581

Cell contents: correlation, *p*-value

\*Correlation significantly different from 0 at the 0.05 level

Argument; PS for Problem Solving; MOD for Modelling; COM for Communication; REP for Representation; SYM for Symbols and Formalism. The prefix 'AVG' indicates the average code value across the eight coders on the relevant competency.

### 2.3.1.2 Coder Consistency

Coder consistency can be approached from two perspectives. The first is the degree to which coders' actual coding of the items correlated with codings they had initially given the items in another coding of the same items 2-years previously. This would be an analysis of intra-coder consistency. The other examination of coder consistency would be an examination of the degree to which the eight coders tended to code in common for a given item relative to the competencies. Such consistency would be an example of inter-coder consistency.

**Table 2.3** Intra-coder consistency for common PISA 2003/PISA 2006 items

Coder/competency	SYM	REA	PS	MOD	COM	REP
Coder 1	0.804*	0.803*	0.805*	0.885*	0.847*	0.860*
Coder 2	0.644*	0.906*	0.777*	0.856*	0.855*	0.884*
Coder 3	0.505*	0.575*	0.459*	0.380*	0.652*	0.703*
Coder 4	0.369*	0.428*	0.438*	0.579*	0.462*	0.404*

\* $r$  is significantly different from 0 at the 0.05 level

*Intra-coder consistency.* Within-coder data only exist for four of the eight coders in our sample. In addition, there have been minor changes in the description of the competency-related codes for some of the variables that may have slightly altered the use of the codes during the intervening 2 years. These cautions notwithstanding, correlations were conducted for the six competency-related variables for each of the coders for whom there was complete data for the two separate codings of the 48 items. The results are in Table 2.3 Note that all of the observed correlations were significantly different from 0 at the 0.05 level.

There was an interesting pattern in the intra-coder correlations of the coders' work. The coders are numbered in ascending order from most consistent to least consistent. This ordering also matches the ordering of amount of experience and coding the four coders had with using the MEG Item Difficulty Framework. This suggests, perhaps, that coders become more consistent with increased familiarity with the framework and its use, and that training in the use of the framework will be an important issue for the future.

*Inter-coder consistency.* A second approach to coder consistency lies in examining the degree to which the eight coders actually give the same code to an item for a given competency. In essence, this is asking to what degree the eight individual MEG coders give the same numerical code to an item for any one of the six competency-related variables. This analysis can be approached from a variety of perspectives. Historically, most researchers have been satisfied with finding the Pearson product moment correlation of the coders over the set of items related to a given competency area. More recently, researchers dealing with content coding and curricular studies have shifted toward the use of Cronbach's  $\alpha$  along with more emphasis on individual item and coder patterns of behaviour (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Shrout & Fleiss, 1979; von Eye & Mun, 2005).

The data showing the codes each of the eight individual coders awarded for each item have been collected and analysed to determine the consistency of the eight coders for each of the six competency-based variables. Table 2.4 contains a variety of information points for each item. In addition to Cronbach's  $\alpha$  value for each competency-based variable, data are provided showing the distribution of ranges between minimum and maximum codes given to individual items for the competency variable in the coding. Note that a range of 3 for an individual item indicates that it was coded as being at both Level 0 and Level 3 by different coders.

The examination of the values of Cronbach's  $\alpha$  for the six competency-based variables shows considerable consistency with the exception of the Reasoning

**Table 2.4** Inter-coder consistency data for the six competency-based variables

	SYM	REA	PS	MOD	COM	REP
Cronbach's $\alpha$	0.89	0.62	0.90	0.81	0.95	0.83
Average code	1.42	1.52	1.54	1.48	1.58	1.40
Code range=0	3	0	1	1	0	3
Code range=1	26	25	24	24	24	27
Code range=2	15	21	19	22	20	14
Code range=3	4	2	4	1	4	4
Coding outliers	C5=2 C8=2	C2=1 C5=1	C8=4	C6=1	C1=1 C4=1 C5=1 C8=1	C5=2 C8=2

and argumentation variable. An examination of the individual item and response data in general did not immediately indicate a reason for the lower consistency value observed.

The next row in the table indicates the average coding value given across each of the 48 items in each of the competency-based variables' actual coding for this study.

The remaining rows of the table provide a great deal of information about how the 48 items were coded within each of the competency-based variables. The term "Code Range" in the left hand column refers to the range of codes, which is the value of maximum code awarded minus the value of the minimum code. For example a code range of 0 would indicate that all eight coders agreed on the code awarded an item. A code range of 1 would indicate that all coders were awarding one of two adjacent codes. A code range of 3, however, would indicate that at least one coder had awarded a code of 0 while another coder had awarded a code of 3 to an item. Items for which this occurred were flagged for extra analysis. The coders were numbered C1–C8 and the final row in the table indicates which coders were "outliers" in the coding of the individual item receiving a code range of 3.

## 2.3.2 Results of Difficulty Analyses

### 2.3.2.1 Predicting Variance Explained

The degree to which the six competency related variables add to the explanation of variance in the item difficulty scores associated with student performance for the PISA 2003 and PISA 2006 mathematics surveys was analysed using first the "best subsets" approach, and then through a separate multivariate regression analysis of the data.

Best Subset Regressions: Analysis of the PISA 2003 data, the implementation of the "best subset" regression approach, which is sometimes called the "all possible regressions" approach, resulted in the information shown in Table 2.5 (Chatterjee & Price, 1977; Draper & Smith, 1966).



**Table 2.5** Best subset analysis of the PISA 2003 item logit values using the average competency variable scores as predictors

Vars	R-Sq	Adj. R-Sq	C-p	s	AVGSYM	AVGREA	AVGFS	AVGMOD	AVGCOM	AVGREP
1	48.1	47.0	38.6	0.94742		X				
2	68.8	67.4	7.7	0.74293	X	X				
3	72.4	70.5	4.0	0.70669	X	X	X			
4	73.8	71.3	3.7	0.69643	X	X	X		X	
5	74.0	70.9	5.4	0.70206	X	X	X	X	X	
6	74.2	70.5	7.0	0.70705	X	X	X	X	X	X

The values in each row show data associated with various sets of predictor variables and the percentage of variance in the PISA 2003 item difficulty variability they predict. This percentage is given by the value in the Adjusted R-square column. The data show the best predictor variables as the number of variables used in the model goes from 1 to 6.

An examination of the table provides a number of observations. First, the percent of variance predicted (Adjusted R-Squared value) increases up to a four-variable model and then decreases slightly thereafter for the best models with five or six predictor variables. The degree to which the increased value of prediction increases will be discussed later.

It is interesting that the one best competency predictor is the Reasoning and argumentation variable. The entrance of additional predictor variables in building best models with more variables show the entry order of Symbols and formalism, Problem solving, Communication, Modelling, and Representation. The latter two variables do not appear to add to the explanatory power achieved using only the first four.

Table 2.6 shows the same analysis conducted using the PISA 2006 item difficulty estimates. The results are very similar in that the four-variable model appears the best in numerical value and the first four variables entering are the same: Reasoning and argumentation, Symbols and formalism, Problem solving, and Communication. However, there is a slight difference in the order of the entrance of the remaining two variables into the predictor models. Here the next is Representation, followed then by Modelling. However, the data suggest that the addition of these latter two variables does not improve the prediction based on the four variables common to both the PISA 2003 and PISA 2006 data.

Overall, these best subset regression analyses indicate that the four variables of Reasoning and argumentation, Symbols and formalism, Problem solving, and Communication provide the best structure for maximizing the prediction of item difficulty in PISA as defined by item logit values. Additional analysis of the relative contributions of each of these will appear in the next analyses.

Multiple regressions: Table 2.7 contains the results of a stepwise regression employing all possible competency-based variables for the prediction of the PISA 2003 item difficulty logit values. The algorithm was structured to select the best single predictor, and then add the next best single predictor that would add a significant amount of explanatory power. This process iterates, adding variables to the regression equation until the point when the addition of any other variable to the regression equation would no longer make a statistically significant increase in the amount of item difficulty variance explained.

This regression equation indicates that the three competency-based variables, in order of explanatory power are Reasoning and argumentation, Symbols and formalism, and Problem solving. This model predicts 70.5% of the variability in the PISA item difficulty data, when the R-squared value is adjusted. While the addition of the variable Communication would have pushed the R-squared value to 71.8, the gain would not have been statistically significant over the variance explained by this three-variable model.

**Table 2.6** Best subset analysis of the PISA 2006 item logit values using the average competency variable scores as predictors

Vars	R-Sq	Adj. R-Sq	C-p	s	AVGSYM	AVGREA	AVGFS	AVGMOD	AVGCOM	AVGREP
1	48.5	47.4	40.5	0.93106		X				
2	69.3	68.0	8.4	0.72646	X	X				
3	73.2	71.4	4.0	0.68674	X	X	X			
4	74.8	72.4	3.4	0.67423	X	X	X		X	
5	75.0	72.0	5.1	0.67938	X	X	X	X	X	
6	75.0	71.4	7.0	0.68676	X	X	X	X	X	X

**Table 2.7** Stepwise regression for the explanation of variability in the PISA 2003 item difficulty logit values

Step	1	2	3
Constant	-2.212	-2.524	-2.573
AVGREA	1.64	1.32	0.87
T-value	6.53	6.44	3.20
AVGSYM		1.09	1.02
T-value		5.46	5.33
AVGPS			0.67
T-value			2.39
S	0.947	0.743	0.707
R-Sq	48.09	68.78	72.38

The resulting regression equation is:  
 $PISA_{2003} = -2.573 + 0.87 * AVGREA + 1.02 * AVGSYM + 0.67 * AVGPS$

**Table 2.8** Stepwise regression for the explanation of variability in the PISA 2006 item difficulty logit values

Step	1	2	3
Constant	-2.212	-2.521	-2.572
AVGREA	1.62	1.31	0.85
T-value	6.59	6.53	3.22
AVGSYM		1.08	1.01
T-value		5.53	5.42
AVGPS			0.68
T-value			2.52
S	0.931	0.726	0.687
R-Sq	48.53	69.35	73.22

The resulting regression equation is:  $PISA_{2003} = -2.572 + 0.85 * AVGREA + 1.01 * AVGSYM + 0.68 * AVGPS$

Carrying out the same stepwise regression approach using the PISA 2006 item difficulty logit data as the dependent variable, we obtain the results shown in Table 2.8. As in the case of the PISA 2003 data, the same three variables, Reasoning and argumentation, Symbols and formalism, and Problem solving enter in the same order. In this case, the three variables explain 71.4% of the variability in the item difficulty values when the adjusted R-squared value is computed.

A comparison of the coefficients show that there is no difference between the models developed from the PISA 2003 and PISA 2006 data. In like manner, there is no difference in the ascending order in which the three statistically significant predictor variables enter into the equations. In both cases, the calculation of the Durbin-Watson statistic and other residual diagnostics indicate that these models are sound and free of common biasing factors sometimes found in regression model building.

### 2.3.2.2 Factor Analysis

A factor analysis was conducted to examine the structure of the space spanned by the six competency-based variables. A principal components factor analysis of the

**Table 2.9** Factor analysis of the competency-related variable codings

Sorted unrotated factor loadings and communalities						
Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
AVGREA	-0.833	-0.229	-0.355	-0.213	-0.129	0.258
AVGMOD	-0.762	0.181	0.450	0.064	-0.418	-0.075
AVGPS	-0.736	-0.430	0.020	-0.441	0.188	-0.207
AVGSYM	-0.666	0.538	0.353	0.022	0.361	0.110
AVGCOM	-0.554	0.484	-0.605	0.267	-0.002	-0.148
AVGREP	-0.438	-0.588	0.088	0.665	0.112	0.008
Variance	2.758	1.140	0.826	0.758	0.360	0.149
%Var	0.460	0.190	0.138	0.126	0.062	0.025

correlation matrix of the competency-based variable scores for the 48 items revealed the findings shown in Table 2.9. An examination of the data indicates that there were two factors having eigenvalues greater than one. Given that each variable contributes a value of 1 to the eigenvalues total, only the factors having eigenvalues in the end greater than one are considered significant and retained for further study.

An examination of the percent of variance described by the first two factors show that they account for a total of 64% of the variance in the codings. Factor 1's strongest loadings are Reasoning and argumentation, Modelling, Problem solving, and Symbols and formalism. This might be considered, given the values, a balanced factor similar to a generalised academic demand factor. Factor 2's strongest loadings are Symbols and formalism decreased by Representation and Problem solving. This second factor might be considered as describing increased item demand related to the requirement to decode and deal with Symbols and formalism and Communication in the absence of Problem solving and the demand to interpret and manipulate Representations. One might liken this to adding demand for reading and symbol manipulation as it occurs without enacting problem solving strategies or multiple representations of mathematical concepts or operations.

An important remark: It might seem that a certain subset of those six competencies will already serve all purposes and that the others are unnecessary. However, a subset of competencies proved to be sufficient only for explaining item difficulty and only in the particular case of PISA tests. In other cases, other subsets might have more explanatory power. More importantly, the competencies serve a much broader purpose than only explaining item difficulty. For the most important purpose, that is describing proper mathematical activities and thus formulating the essential aims that students ought to achieve through school mathematics, all competencies are indispensable.

## 2.4 Present Status of the Study

The foregoing data provide sufficiently strong evidence of the role played by the mathematical competencies, as defined in Table 2.1, in influencing variability in item difficulty on the PISA mathematics survey items. At present, illustrations of

the way the competencies play out to influence difficulty in particular items are being developed, along with an elaborated coding manual for researchers who have not been involved in the development of the MEG model. This coding manual will be central in the next stage of the study, as it will be used with researchers unfamiliar with it and the coding of PISA items, but familiar with coding structures. They will be asked to code the 48 PISA items and their results will be compared with those of the MEG members.

The planned next steps are as follows. Based on this experience and revisions that may result from observing these coders and their work, a broader field test shall be conducted where new individuals, familiar with the PISA project, will be asked to use the coding instruction manual without any other assistance to code the 48 items. Their coding results and written comments shall again be used to further the development of the model and manual for either one more round of field testing or release as a PISA technical report.

Two further developments of this study might be to investigate the extent to which the scheme could be used to predict the difficulty of newly developed PISA mathematics items; and to investigate its applicability to other (non-PISA) mathematics items.

Curriculum statements in many countries reflect the importance of the competencies on which this study has focused. It can be expected that the relationship between cognitive demand for the activation of these competencies and the empirical difficulty of the mathematical tasks that call for such activation, whether in the PISA context or in other contexts, will be of deep interest to teachers, teacher educators and others involved in mathematics education around the world.

## References

- Blum, W., Druke-Noe, C., Hartung, R., & Köller, O. (2006). *Bildungsstandards Mathematik: konkret*. Berlin: Cornelsen-Scriptor.
- Chatterjee, S., & Price, B. (1977). *Regression Analysis by Example*. New York: Wiley.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.
- De Lange, J. (1987). *Mathematics, Insight and Meaning*. Utrecht: CD-Press.
- Draper, N. R., & Smith, H. (1966). *Applied Regression Analysis*. New York: Wiley.
- Neubrand, M., Biehler, R., Blum, W., Cohors-Fresenborg, E., Flade, L., Knoche, N., et al. (2001). Grundlagen der Ergänzung des internationalen OECD/PISA-Mathematik-Tests in der deutschen Zusatzerhebung. *Zentralblatt für Didaktik der Mathematik*, 33(2), 45–59.
- Niss, M. (1999). Kompetencer og uddannelsesbeskrivelse. *Uddannelse*, 9, 21–29.
- Niss, M. (2003). Mathematical Competencies and the Learning of Mathematics: The Danish KOM Project. In A. Gagatsis & S. Papastavridis (Eds.), *3rd Mediterranean Conference on Mathematical Education* (pp. 115–124). Athens: The Hellenic Mathematical Society.
- Niss, M., & Hoejgaard, T. (eds.) (2011) *Competencies and mathematical learning. Ideas and inspiration for the development of mathematics teaching and learning in Denmark*. English edition, October 2011 (IMFUFAttekst no. 485). Roskilde: Roskilde University.
- OECD. (2003). *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*. Paris: Directorate for Education, OECD.

- OECD. (2004). *Learning for Tomorrow's World – First Results from PISA 2003*. Paris: Directorate for Education, OECD.
- OECD. (2006). *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*. Paris: Directorate for Education, OECD.
- Planas, N. (2010). *Pensar i comunicar matemàtiques*. Barcelona: Fundació Prepedagògic.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin*, 86(2), 420–428.
- Von Eye, A., & Mun, E. Y. (2005). *Analyzing rater agreement: manifest variable methods*. Mahwah: Lawrence Erlbaum. [www.ug.dk/uddannelser/professionsbacheloruddannelse/enkeltfag](http://www.ug.dk/uddannelser/professionsbacheloruddannelse/enkeltfag). Visited 19 May 2011.