OECD Education Working Papers No. 87

Analysis of the Predictive Power of PISA Test Items

Maciej Jakubowski

https://dx.doi.org/10.1787/5k4bx47268g5-en

OECD

Unclassified **EDU/WKP(2013)2**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

English - Or. English

DIRECTORATE FOR EDUCATION

EDU/WKP(2013)2
Unclassified

Cancels & replaces the same document of 31 January 2013

ANALYSIS OF THE PREDICTIVE POWER OF PISA TEST ITEMS

OECD Education Working Paper no. 87

This working paper is prepared by Maciej Jakubowski (Warsaw University, Poland). It is based on the paper developed by Kaye Forgione (Achieve, United States) and the analyses conducted by Artur Pokropek (Educational Research Institute IBE, Poland).

For further information, please contact Miyako Ikeda, Miyako.IKEDA@oecd.org

English - Or. English

OECD DIRECTORATE FOR EDUCATION

OECD EDUCATION WORKING PAPERS SERIES

This series is designed to make available to a wider readership selected studies drawing on the work of the OECD Directorate for Education. Authorship is usually collective, but principal writers are named. The papers are generally available only in their original language (English or French) with a short summary available in the other.

Comment on the series is welcome, and should be sent to edu.contact@oecd.org.

---

www.oecd.org/edu/workingpapers

TABLE OF CONTENTS

ABSTRACT

The predictive power of the PISA test items for future student success is examined based on data from the Longitudinal Surveys of Australian Youth (LSAY) for the PISA 2003 cohort. This working paper analyses how students' responses to mathematics and problem-solving items in PISA 2003 are related to the students' qualifications in education in 2007 and 2010. The results show that items do differ in their predictive power, depending on some of their deep qualities. PISA mathematics and problem-solving items are grouped into various classifications according to their qualities. This paper proposes 16 new classifications of items. Among mathematics-specific item classifications, two are found to be significantly related to future student success: those that assess knowledge, understanding, and application of statistics; and those related to rates, ratios, proportions, and/or percent. These items frequently require students to apply common mathematical concepts to solve multi-step, non-routine problems, think flexibly, and understand and interpret information presented in an unfamiliar format or context. Among classifications that are not specific to mathematics, items that were classified as using reverse or flexible thinking are found to be related to student qualifications in both mathematics and problem solving. These items require students to be able to think through a solution at various points during the problem-solving process, not just at the start.

RÉSUMÉ

L'efficacité prédictive des items de l'enquête PISA pour la réussite future des élèves est examinée à partir de données collectées par l'enquête longitudinale australienne LSAY (Longitudinal Surveys of Australian Youth) sur l'échantillon d'élèves évalués lors du cycle PISA 2003. Le présent document de travail analyse dans quelle mesure les réponses des élèves aux items de mathématiques et de résolution de problèmes de l'enquête PISA 2003 sont liées à leur niveau de formation en 2007 et 2010. Les résultats montrent que l'efficacité prédictive des items varie en fonction de certaines de leurs qualités profondes. Les items PISA de mathématiques et de résolution de problèmes sont classés dans différentes catégories selon leurs qualités. Le présent document propose 16 nouvelles catégories d'items. Parmi les catégories d'items spécifiques à l'évaluation des compétences en mathématiques, deux ont été identifiées comme étant liées de façon significative à la réussite future des élèves : les types d'items qui évaluent les connaissances, la compréhension et l'application des statistiques, et ceux qui ont trait aux taux, ratios, proportions et/ou pourcentages. Ces items requièrent généralement des élèves l'application de concepts mathématiques de base pour résoudre des problèmes non routiniers et comportant plusieurs étapes, un raisonnement flexible, ainsi que la compréhension et l'interprétation d'informations présentées sous un format ou dans un contexte non familiers. Parmi les catégories qui ne sont pas spécifiques aux mathématiques, il apparaît que les items définis comme faisant appel à un raisonnement rétrospectif ou flexible présentent une corrélation avec les qualifications des élèves à la fois en mathématiques et en résolution de problèmes. Ces items supposent de la part des élèves une capacité à réfléchir à une solution à différentes étapes du processus de résolution du problème, et pas uniquement au début de ce processus.

ANALYSIS OF THE PREDICTIVE POWER OF PISA TEST ITEMS

I. Introduction

1. Several countries and economies that have participated in PISA have also established longitudinal studies based on sampled cohorts of 15-year-olds tested in PISA. Longitudinal follow-up studies of students who participated in PISA offer the possibility of exploring how competencies at age 15 are related to educational careers and labour-market outcomes. The richest longitudinal data available come from the PISA 2000 participating cohorts in Canada (YITS) and Switzerland (TREE) and from two Australian studies (LSAY) that are based on the PISA 2003 and 2006 cohorts. All these surveys collect information on students' educational careers after the age of 15, which is highly associated with future labour-market outcomes. The YITS and TREE studies also collect information on early labour-market outcomes, but preliminary analysis shows that because of the young age of the individuals sampled in these longitudinal studies, most of the highly skilled participants are still in education.

2. This report summarises the project that aimed to classify PISA test items into different groups according to what they can predict about future outcomes, based on the LSAY data for the PISA 2003 cohort. Analysis shows that items do differ in their predictive power and these differences are not only related to cognitive test characteristics of these items (*i.e.* difficulty or discriminatory power), but also to some deeper qualities. The report proposes several classifications of PISA mathematics and problem-solving items that were hypothesised to be related to future outcomes. Statistical analysis presented in the report shows which classifications characterise items that in fact have higher predictive power.

3. Several classifications proposed can be generalised to other domains, not only mathematics or problem solving. These classifications define deeper item qualities that can be common across school subjects. Among these, items that were classified as those *using reverse or flexible thinking* were found to be related to future student outcomes. These items require students to be able to think through a solution from points in the solution process other than the start.

4. Among mathematics-specific item classifications, two were determined to be significantly related to future student success. Items that assess knowledge, understanding, and application of *statistics* and items that assess knowledge, understanding, and application of *rates, ratios, proportions, and/or percent* appear to demonstrate predictive power. Close examination of the items coded to these two classifications reveals that they frequently call for students to apply common mathematical concepts to the solution of multi-step, non-routine problems, to think flexibly, and to understand and interpret information presented in an unfamiliar format or context.

5. The report is structured as follows. The next section describes the data and methods used in this study. The third section briefly discusses newly proposed and original classifications of PISA items. The fourth section presents the results of the statistical analysis relating items classifications to later student outcomes. This section also discusses results for single items in mathematics and problem solving tested in PISA 2003 and how well items of different formats measure the skills related to later student outcomes. The final section summarises the findings.

II. Data and methods

PISA cognitive data

6. The main results of PISA are presented on scales summarising students' achievement in different assessment areas. Each PISA cycle devotes more testing time to one assessment area. PISA also reports results on subscales that measure achievement in subdomains. In PISA 2003 the main domain was mathematics. PISA reports not only achievement on the main scale but also on four mathematics subscales.

7. Besides the scaled achievement data, PISA publishes student responses to individual test items. These data are freely available and can be used for any type of re-analysis of performance on PISA. Although the item-level data are reliable and carefully coded, because the booklets are rotated, not all students answer the same single items, and sample sizes are much smaller. Also, items are differently positioned in test booklets; that might affect student performance on a particular item, especially if an item is at the end of the booklet.[1] These features of the assessment design need to be taken into account when interpreting results based on single PISA items.

Longitudinal Surveys of Australian Youth (LSAY)

8. The Longitudinal Surveys of Australian Youth (LSAY) track young people as they move from school into further study, work and other destinations. It uses large, nationally representative samples of young people to collect information about education and training, work and social development. Basic information about the programme, and basic results, micro-data and reports, are all available on the programme's website: http://www.lsay.edu.au/. The datasets and their descriptions are based on the information provided there.

9. Since 2003, the survey has been integrated with the PISA study. After the PISA 2003 survey was conducted, 15-year-old students who were sampled in Australia were contacted again to establish a first wave of the LSAY/PISA 2003 study. Out of 12 551 students in the original Australian PISA sample, 10 370 students were contacted and found eligible for the LSAY study. Table 1 shows the response rates for all waves of the 2003 study. The results presented in this paper are mostly based on the 2007 and 2010 waves of the LSAY/PISA 2003 study.

Table 1. Response rates for the LSAY/PISA 2003 study from 2003 to 2010

Year	2003	2004	2005	2006	2007	2008	2009	2010
Wave	1	2	3	4	5	6	7	8
Average age	15.7	16.7	17.7	18.7	19.7	20.7	21.7	22.7
No. of respondents	10 370	9 378	8 691	7 721	6 658	6 074	5 475	4 903
% of Wave 1	100	90.4	83.8	74.5	64.2	58.6	52.8	47.3

Source: http://www.lsay.edu.au/cohort/2003/104.html

1 . In PISA 2003, the 167 main study items were allocated to 13 item clusters (seven mathematics clusters and two clusters in each of the other domains), with each cluster representing 30 minutes of test time. Each of 13 test booklets contains four clusters. Each cluster appears in each of the four possible positions within a booklet exactly once. When estimating the item parameters, booklet effects were included in the measurement model to prevent confounding item difficulty and booklet effects.

10. LSAY datasets provide all data available in the international PISA datasets. In addition, students are contacted every year to collect information on school and post-school activities, including work experience and what students do when they leave school. This includes vocational and higher education, employment, job-seeking activities, and satisfaction with various aspects of their lives.

11. In this study, several indicators of post-school student outcomes are examined. As students in the LSAY/PISA 2003 study are younger than the age at which all of them should have gained labour-market experience, only outcomes describing post-secondary school achievement provide comparable information on student success. From those, this report mainly focuses on the derived indicator of obtained or current qualification level. The indicator of qualification level was derived from two indicators: (*i*) the current qualification level reported at the time of the interview and (*ii*) the highest qualification completed at the time of the interview. Both indicators have 10 distinct values, which are presented in Table 2. As shown in the table, these values were re-coded into five categories, giving a reasonable number of observations in each category and providing the strongest linear relationship with PISA performance in mathematics. The indicator was computed by taking the highest qualification level completed at the time of the interview and replacing missing values on this indicator with values of the current qualification level to increase the number of non-missing observations. For the highest qualification level completed, the last category "Did not complete a qualification" was replaced with a missing value unless there was no missing information on the current qualification level reported, in which case the current level was used.

Table 2. Re-coding of obtained or current qualification level for the LSAY/PISA 2003 study

qualification level	New outcome variable after recoding	Highest qualification level completed (XHEL)				Current qualification level (XCEL)			
		% in population		% of those with some qualifications		% in population		% of those with some qualifications	
		2007	2010	2007	2010	2007	2010	2007	2010
1 Certificate I	1	1.4	3.4	6.5	5.7	2.8	0.8	5	2.4
2 Certificate II	2	3.3	4.7	16	7.9	1.7	0.6	2.9	1.9
3 Certificate III	2	6.5	11.6	31.3	19.6	6.5	3.4	11.6	10.3
4 Certificate IV	3	2.8	5.6	13.6	9.3	2.1	2.2	3.8	6.6
5 Certificate - level unknown	3	1	2.2	4.5	3.7	2.9	1.3	5.1	3.8
6 Advanced diploma/diploma (incl. associate degree)	3	4.8	8.2	23	13.8	4.7	3.2	8.3	9.7
7 Bachelor degree	4	0.9	22.4	4.5	37.8	35.3	18.6	62.8	56.1
8 Graduate diploma/graduate certificate	4	0.1	0.5	0.5	0.8	0.2	0.4	0.3	1.3
9 Postgraduate degree (PhD/Masters)	5	0	0.9	0	1.4	0.1	2.7	0.2	8
XHEL: 10 Did not complete a qualification XCEL: 10 Not studying for a qualification	Missing	79.2	40.6	-	-	43.8	66.8	-	-

Methods

12. The predictive power of PISA items is assessed by explaining differences in student qualification levels in 2007 and 2010 using students' responses to items in 2003. The relationship is examined for both single items and scales based on sets of items. The main results for mathematics and problem-solving items presented in the paper were obtained using a linear regression model, which is applied to individual

student-level data from the LSAY/PISA 2003 study. Three sets of regressions are estimated. The first set relates later student outcomes to item-based information only. In the second set, the relationships between students' responses to items and later student outcomes are examined after accounting for the overall performance level of a student (using the first plausible value in mathematics or problem solving). In the third set, which is presented in Annex 1, in addition to performance level, the following student characteristics are accounted for: gender, socio-economic background (the ESCS index) and immigrant background (two dummy variables for first- and second-generation students and a dummy denoting students speaking a different language at home than the language of the test).

13. Data were weighted using weights provided by the LSAY research team. These give results representative of the population of students tested in PISA 2003, adjusted for missing responses of those who dropped from the study in the intervening years. Standard errors were calculated using Taylor linearised variance estimation accounting for clustering of students at the school level.

14. Item psychometric characteristics, like item difficulty or item discriminatory power, can be correlated with the predictive power of PISA items. On average, the more difficult an item is, the stronger should be the correlation between the correct response to this item and positive later outcomes. Thus, main results are also presented before and after accounting for item difficulty and item discriminatory power. Item difficulty was measured using the IRT model and data are available in PISA Technical Reports. Item discriminatory power was estimated using the two-parameter IRT model (2PL) for the purpose of this study.

15. Original PISA classifications of items as well as new classifications were used to produce scales based on responses to a class of test items. Although the IRT modelling is a proper technique used in PISA to scale the cognitive data, for the purpose of this study the principal component analysis based on polychoric and polyserial correlations was employed to derive item-based scales. The results obtained with the IRT model would be very similar. The estimated scales were standardised as z-scores with weighted mean 0 and standard deviation 1 in the original PISA 2003 sample.[2]

III. Classifications of PISA items

16. For the purpose of this study, the OECD, in co-operation with Achieve (http://www.achieve.org/), launched a study to propose new classifications of PISA items. The aim of this study was to classify mathematics items from PISA 2000 and PISA 2003 and problem-solving items from PISA 2003 according to attributes that Achieve's mathematics experts hypothesised might help define and describe differences in predictive power as it relates to post-secondary success. A team of three experts in mathematics and/or mathematics education reviewed all of the mathematics items from PISA 2000 and PISA 2003 and the problem-solving items from PISA 2003. They originally proposed a set of more than 20 classifications that they hypothesised might relate to the degree of predictive power of an item and coded each item for the presence or absence of these attributes. However, some of the attributes were represented in only a few items and they were not considered for further statistical analysis.

17. The list below provides names of newly developed classifications that are considered in this study.

2. Caution is advised in drawing inferences based on these indices. Further analysis is required to understand the influence of measurement error (i.e. reliability of scales) on the parameters estimated.

1. Thinking across multiple representations – Items that require students to demonstrate the ability to use, apply, or translate between multiple representations were coded as having this attribute present. Items might call on students to translate between such varying mathematical representations as tables, graphs, narrative descriptions, and/or pictorial representations. They might also require students to use information from multiple representations to solve contextualised problems.

2. Probabilistic reasoning – Items that require probabilistic reasoning address such challenging and often misunderstood concepts as likelihood and prediction. While an item may not necessarily explicitly refer to probability, the notion may nonetheless be present in an item; if this was the case, an item was coded as having this attribute.

3. Statistics – Items that assess one or more statistical concepts and procedures were coded as having this attribute present. Such concepts included averages and other measures of centre, measures of spread, statistical outliers, and variance.

4. Rates, ratios, proportions, and percentages – Rates, ratios, proportions, and percentages tend to be challenging but important concepts for 15-year-old students and also important concepts for solving real-life problems. Items were coded as having this attribute if they explicitly require the use of one or more of these concepts OR if one of these concepts is the basis of the task, even if it is not explicitly required for solving the problem.

5. Measurement – Items the expert reviewers found were related to measurement skills or concepts addressed a range of topics, including the use of a formula to calculate such constructs as area, perimeter, volume, circumference, or time. The preponderance of items coded as possessing this attribute involved geometric measurement. The notion of measurement units was not an issue since solution options either provided units as part of the list of possible answers or the coding guides made clear that whether students provided correct units was not a factor in awarding students full credit.

6. Using graphs, charts, tables, pictures, and diagrams – A large number of PISA items require students to read or interpret some sort of graphic organiser or display. Such depictions include graphs, charts, tables, pictures, and diagrams. While the notions of both reading and interpreting are captured by this attribute, separate codes are assigned for "no display" (=0), "reading a display" (=1) and "interpreting a display" (=2).

7. Multiple graphs, charts, tables, pictures, and/or diagrams – A number of PISA items require students to use information provided in more than one graphic display (*i.e.* graphs, charts, tables, pictures, and diagrams). Items were coded to indicate whether they include no graphic organisers (=0), one graphic organiser (=1), or multiple graphic organisers (=2). The expert reviewers hypothesised that this might be an important factor connected with items' predictive power, so they chose to track this attribute separately.

8. Spatial reasoning – The expert reviewers posited that students' success with items requiring spatial visualisation and reasoning might be connected with their predictive power. Items identified as possessing this attribute might require students to visualise the rotation of an object or how it might look from different perspectives.

9. Conceptual understanding vs. procedural knowledge – The National Research Council's Adding It Up: Helping Children Learn Mathematics (2001) defines conceptual understanding and procedural fluency as two of the five components of mathematical proficiency. In this seminal

publication, conceptual understanding is defined as the "comprehension of mathematical concepts, operations, and relations." Test items that assess conceptual understanding require students to exhibit "an integrated and functional grasp of mathematical ideas." Procedural knowledge is a component of procedural fluency, as defined in Adding It Up. Achieve's expert reviewers purposefully identified procedural knowledge, rather than procedural fluency as the attribute appropriate for this study since the notion of fluency (flexible, accurate, and efficient application of mathematical procedures) is not clearly assessed through PISA. Procedural knowledge is, however, assessed in PISA. For the purposes of this study, it is defined as the knowledge of procedures (*e.g.* algorithms, rules, plans) and the knowledge of when and how to use these procedures appropriately – both of which are components of the definition for procedural fluency provided in Adding It Up.

10. Constructing or presenting an argument or explanation – Items deemed to have this attribute go beyond requiring students to show their work. Rather, items possessing this attribute require students to construct or present an argument, explanation, or justification.

11. Systematic, sequential, or strategic reasoning – A substantial number of PISA items require students to display complex thinking processes that can best be defined as systematic, sequential, and/or strategic. By using the term "systematic" thinking, the expert reviewers identified those items that require students to analyse or create complex plans or diagrams that function as a whole. "Sequential reasoning" is involved in items that require students not simply to follow directions, but to create a sequence that needs to be followed in a particular order. "Strategic reasoning" describes those items that call for students to use such strategies as guess and check, drawing a diagram, looking for a pattern, making a list, or solving a simpler, related problem. Items that involve scheduling or that require students to use counting techniques, such as combinatorics or tree diagrams, were coded as having this attribute present.

12. Applying information/learning through studying examples – While a limited number of PISA items require students to learn through the study of examples and then apply that learning to answer questions, the expert reviewers surmised that this is a skill that is important for independent learners – hence for success in post-secondary endeavours.

13. Thinking through a multi-step solution process – These items involve several steps, and although they are well-defined, they are not practiced or rote. Items might include irrelevant data. Although complex multiple-choice items may require several one-step processes, they are not necessarily coded as multi-step.

14. Employing conceptual understanding/reasoning – These items require students to do one or more of the following:

- Show understanding of commonly held misconceptions of a concept;
- Reflect on whether the results make sense;
- Create or choose an explanation;
- Attend to the meaning of results, not just how to attain them;
- Analyse relationships to draw conclusions;
- Draw conclusions from observations, citing evidence;

- Develop a logical argument that supports a concept or position;
- Use information or concepts to solve problems or respond to questions;
- Reason, plan, and use evidence that requires a higher level of thinking (sometimes abstract); and/or
- Make conjectures.

15. Using reverse thinking/flexible thinking – These items require that students be able to think through a solution from points in the solution process other than the start:

- Given the result, determine how that result may have been derived.
- Determine how a change in input affects either the output or solution path.
- Interpret results in the context of the situation.
- Know and flexibly use properties or representations.
- Analyse solution choices and relate them back to the stimulus/prompt.

16. Making sense of non-routine situations that could require a strategic solution path –These items tend to encourage logical thinking, and allow for a transfer of domain-specific strategies to resolve unfamiliar situations. An item might:

- Require students to use a strategic solution process, rather than a practiced or rote process;
- Require students to make sense of the data and the information provided in order to design a solution strategy;
- Require sense-making;
- Focus on a higher level of interpretation and organisation; and/or
- Involve an unfamiliar context.

18. The above classifications of 1 to 5, 8, and 10 to 12 were coded for all items as 0 or 1, denoting items that belong to each attribute. The classifications of 6 listed above (*using graphs, charts, tables, pictures, and diagrams*) was coded as on a three-step scale (2 = interpreting a display; 1 = reading a display; and 0 = no display). The classifications of 7 (*multiple graphs, charts, tables, pictures, and/or diagrams*) were coded as on a three-step scale (2 = multiple graphic organisers; 1 = one graphic organiser; 0 = no graphic organisers). The classification of 9 (*conceptual understanding vs. procedural knowledge*) was coded as -1/1 where -1 was for procedural and 1 was for conceptual. The last four classifications of 13 to 16 (*thinking through a multi-step solution process; employing conceptual understanding/reasoning; using reverse thinking/flexible thinking; making sense of non-routine situations*) were coded as on a three-step scale (2 = contributes to the cognitive load/demand for an item; 1 = present in an item, but does not determine its cognitive load/demand; 0 = does not contribute to the cognitive load/demand of an item). The first eight classifications of 1 to 8 are specific to mathematics, the last eight classifications of 9 to 16 can be generalised to other assessment areas, such as reading or science.

Original PISA classifications

19. The PISA assessment framework also proposed several different classifications of PISA items. Those are analysed in the report and are related to future student outcomes. The classifications presented in Table A1 are described in details in the PISA 2003 assessment framework (OECD, 2003). Below is a list of classifications used in this study:

- ***Overarching idea***. Each item is classified as representing one overarching idea in mathematics. The PISA 2003 Mathematics Framework defines four such overarching ideas for mathematical content: quantity, space and shape, change and relationships, and uncertainty (see OECD, 2003, pp.34-37). It is interesting to note that the PISA 2000 mathematics items examined are assigned to only two overarching ideas: space and shape, and growth and change.

- ***Competency clusters***.[3] PISA chose to describe the cognitive activities that these competencies encompass according to three *competency clusters*: the *reproduction* cluster, the *connections* cluster, and the *reflection* cluster (for details see OECD, 2003, pp.41-48). Competencies in the "reproduction" cluster involve reproducing practised knowledge. The *connections* cluster competencies build on the *reproduction* cluster skills in taking problem solving to situations that are not routine, but still involve familiar, or quasi-familiar, settings. Competencies in the "reflection" cluster include an element of student reflection about the processes needed or used to solve a problem. They relate to students' abilities to plan solution strategies and implement them in settings that contain more elements and may be more "original" (or unfamiliar) than those in the connections cluster.

- ***Content area***. PISA also classifies items according to their mathematical content or curriculum strands: algebra, discrete mathematics, functions, geometry, number, probabilistic reasoning and statistics.

20. The PISA 2003 Problem Solving Framework defines three problem types: decision making, system analysis and design, and trouble shooting. These three problem types cover most of the problem-solving processes generally identified within the problem-solving domain (see for details OECD, 2003, pp. 160-170). The analysis of problem types has to be limited due to a small number of problem-solving items available for classifications.

Item format

21. The PISA assessment uses items of different formats. This report analyses how skills measured by items of different formats are related to future student outcomes. Five formats were analysed: short response, open-constructed response, closed-constructed response, multiple choice and complex multiple choice.

22. Table 3 presents the list of new mathematics classifications as well as the original PISA classifications and item formats, as well as the number of mathematics items that belong to each classification. Classifications for individual mathematics items are presented in Table A1 and problem-solving items in Table A2 in Annex A.

3 . In PISA 2012, the seven fundamental mathematical capabilities are identified and these are used instead of "competency clusters". The seven fundamental mathematical capabilities used in the PISA 2012 mathematics framework are as follows: communication; mathematising; representation; reasoning and argument; devising strategies for solving problems; using symbolic, formal and technical language and operations; using mathematical tools (OECD, 2013).

Table 3. Various classifications of PISA 2003 mathematics items.

Types of classifications	Items classification (scale)		The number of mathematics items
Newly developed classifications	Thinking across multiple representations		3
	Probabilistic reasoning		5
	Statistics		10
	Rates, ratios, proportions, and percentages		10
	Measurement		10
	Using graphs, charts, tables, pictures, and diagrams	Reading a display	14
		Interpreting a display	16
	Multiple graphs, charts, tables, pictures, and/or diagrams	One graphic organiser	20
		Multiple graphic organisers	10
	Spatial reasoning		12
	Conceptual understanding vs. procedural knowledge	Procedural know ledge	65
		Conceptual understanding	19
	Constructing or presenting an argument or explanation		9
	Systematic, sequential, or strategic reasoning		9
	Applying information/learning through studying examples		1
	Thinking through a multi-step solution process	Present in an item but does not determine its cognitive load/demand	9
		Contributes to the cognitive load/demand for an item	23
	Employing conceptual understanding/reasoning	Present in an item but does not determine its cognitive load/demand	4
		Contributes to the cognitive load/demand for an item	19
	Using reverse thinking/flexible thinking	Present in an item but does not determine its cognitive load/demand	8
		Contributes to the cognitive load/demand for an item	15
	Making Sense of Non-Routine Situations	Present in an item but does not determine its cognitive load/demand	8
		Contributes to the cognitive load/demand for an item	17
Overarching idea	Space and shape		20
	Change and relationship		21
	Uncertainty		21
	Quantity		22
Content area	Algebra		3
	Discrete mathematics		5
	Functions		9
	Geometry		18
	Number		26
	Probabilistic reasoning		5
	Statistics		18
Competencies	Connections		39
	Reflection		19
	Reproduction		26
Item format	Closed Constructed Response		13
	Open Constructed Response		21
	Multiple Choice		16
	Complex Multiple Choice		12
	Short Response		22

IV. Results for mathematics and problem solving

Classifications of items

23. Table 4 shows the results of regression analysis using the Australian longitudinal data based on PISA 2003 (LSAY/PISA 2003). Two sets of results are presented, separately for data collected in 2007, when students were 19 to 20 years old, and the latest available data from 2010, covering students aged 22 to 23.[4] The first set of results show regression coefficients from a model where every scale based on a class of test items was used to explain variation in student qualifications. Clearly, all scales (or classifications) of mathematics achievement are positively related to higher student qualifications. Coefficients show the effect of a one standard deviation change in the scaled item responses on the qualification level.

24. The second set of results shows similar coefficients, but from regression that, in addition to the scaled response to items classified according to an attribute, includes the overall performance in mathematics.[5] These coefficients show how better performance on a particular classification of test items affects student qualifications, in addition to showing student's overall performance. In other words, these are the effects of classifications of items, net a student's overall ability. As students who achieved higher scores in mathematics were more likely to answer individual items correctly in PISA 2003, and these students also tended to attain higher qualification levels in 2007 and 2010, a student's overall performance needs to be accounted for.[6]

25. Among newly developed classifications, item responses classified as *statistics* are shown to be related to student qualifications in 2007 and in 2010. Collectively, *statistics* items all require students to make sense of the information they are given and apply their knowledge and skills in a way that extends beyond the routine application of a well-defined algorithm. Eight of the ten items involve the concept of average or mean but extend beyond the straight-forward application of the algorithm by eliciting more complex reasoning and thinking. For example, two items require students to reason about the effect of an outlier on the mean, and three items require students to determine the mean after a data point has been added or changed. The predictiveness of this set of items appears to lie in the ability of prepared students to be flexible in their thinking and to apply common mathematical concepts to the solution of multi-step, non-routine problems.

26. Item responses classified as *rates, ratios, proportions, and percentages* are shown to be related to student qualifications in 2007 and to a lesser degree in 2010. Ten of the PISA 2003 mathematics items were coded as assessing *rates, ratios, proportions, and percentages*. As with the *statistics* items, all of these items require students to make sense of the information they are given, rather than simply applying a well-defined algorithm to a routine problem. Seven of the items are complex and require multiple steps to arrive at a solution. For example, one open-response item requires students not only to substitute a formula to derive a solution, but to then go through multiple unit conversions before arriving at a correct solution expressed in the appropriate unit. A number of these items also involve contexts that may not be uniformly

4. These dependent variables were not included in the conditioning model for drawing plausible values.

5. These are preliminary analyses using the first plausible value as a proxy for overall student ability when no independent measures of ability are available. Under the Rasch model, the predictive power of students' responses to items after the overall ability level has been accounted for is an indication of Differential Item Functioning (DIF). Caution is advised in drawing inferences based on these results. These analyses will be revisited and improved in the future.

6 . In Tables A3 and A4 in Annex A, the third set of results is also presented. These are the regression coefficients when the student's gender, socio-economic background and immigrant status were taken into account in addition to overall student performance.

familiar to 15-year-old students across a range of countries. Such contexts include the notion of gear ratio, currency exchange rates, population pyramids, and percentage capacity. The predictiveness of this set of items appears to lie in the ability of prepared students to persist through the multiple steps required to solve non-routine problems, and understand and interpret information presented in an unfamiliar format or context.

27. *Using reverse thinking/flexible thinking* is significantly related to student qualifications in 2007. Similarly, positive responses to items belonging to a classification called *constructing or presenting an argument or explanation* is related to student qualifications in 2007, although these relationships are significant only at 10% level, after accounting for overall performance.

28. From the four original PISA performance subscales representing overarching ideas, only the *space and shape* subscale is not related to student qualifications in 2010, after accounting for overall performance. In general, however, all four subscales are positively related to higher student qualifications. In this case, the subscales available in the PISA original dataset were used.

29. From the scales based on content of mathematics items, only content area *number* is significantly related to student qualifications in 2007, after accounting for overall performance.

30. Surprisingly, among the three originally defined subscales of mathematical competencies, the subscale representing items measuring *reproduction* competencies shows the significant relationship with student qualifications in both 2007 and 2010, even after accounting for overall performance in mathematic.

Table 4. Relationship between various mathematic scales in PISA 2003 and student qualifications measured in 2007 and 2010

Types of classifications	Items classification (scale)	2007		2010	
		Simple regression	After accounting for overall performance (1st plausible value in mathematics)	Simple regression	Adjusted for overall performance (1st plausible value in mathematics)
		Coef.	Coef.	Coef.	Coef.
Newly developed classifications	Thinking across multiple representations	0.131***	0.013	0.102***	-0.001
	Probabilistic reasoning	0.183***	0.005	0.182***	0.019
	Statistics	0.283***	0.068***	0.267***	0.077**
	Rates, ratios, proportions, and percentages	0.220***	0.059**	0.194***	0.047*
	Measurement	0.180***	0.014	0.169***	0.021
	Using graphs, charts, tables, pictures, and diagrams	0.309***	0.003	0.272***	0.002
	Multiple graphs, charts, tables, pictures, and/or diagrams	0.309***	0.003	0.272***	0.002
	Spatial reasoning	0.193***	-0.016	0.165***	-0.023
	Conceptual understanding vs. procedural knowledge	0.272***	0.035	0.243***	0.033
	Constructing or presenting an argument or explanation	0.238***	0.047*	0.206***	0.038
	Systematic, sequential, or strategic reasoning	0.245***	-0.001	0.209***	-0.014
	Applying information/learning through studying examples	0.126***	-0.003	0.115***	0.001
	Thinking through a multi-step solution process	0.304***	0.022	0.271***	0.021
	Employing conceptual understanding/reasoning	0.268***	0.030	0.249***	0.040
	Using reverse thinking/flexible thinking	0.325***	0.065**	0.274***	0.041
	Making Sense of Non-Routine Situations	0.265***	0.004	0.228***	-0.005
Overarching idea	Space and shape	0.451***	0.072*	0.409***	0.072
	Change and relationship	0.503***	0.265***	0.460***	0.238***
	Uncertainty	0.514***	0.297***	0.462***	0.242***
	Quantity	0.490***	0.211***	0.445***	0.185***
Content area	Algebra	0.125***	0.036	0.078***	0.006
	Discrete mathematics	0.233***	0.035	0.209***	0.040
	Functions	0.224***	0.022	0.196***	0.010
	Geometry	0.226***	0.000	0.208***	0.011
	Number	0.309***	0.059*	0.282***	0.049
	Probabilistic reasoning	0.183***	0.005	0.182***	0.019
	Statistics	0.267***	0.025	0.244***	0.031
Competencies	Connections	0.312***	0.029	0.285***	0.038
	Reflection	0.236***	0.034	0.186***	0.011
	Reproduction	0.354***	0.066***	0.322***	0.069**

Notes: ***, **, * denotes statistical significance at the 1%, 5% and 10% level, respectively.

31. Similar analyses can be performed for problem-solving items. In this case, in the regression analysis, the overall performance on the problem-solving scale was taken into account when comparing scales that are at least partly based on problem-solving items.

32. The results in Tables 5 suggest that three scales that are based on problem-solving items are positively related to student qualifications: *employing conceptual understanding/reasoning, using reverse thinking/flexible thinking,* and to a lesser extent *thinking through a multi-step solution process.*

Table 5. Relationship between various problem-solving scales in PISA 2003 and student qualifications measured in 2007 and 2010

Types of classifications	Items classification (scale)	2007		2010	
		Simple regression	After accounting for overall performance (1st plausible value in problem solving)	Simple regression	After accounting for overall performance (1st plausible value in problem solving)
		Coef.	Coef.	Coef.	Coef.
Problem types	Decision making	0.119***	-0.034	0.116***	-0.030
	System analysis and design	0.102***	-0.037	0.090***	-0.040*
	Troubleshooting	0.112***	-0.041	0.080***	-0.065**
Newly developed classifications	Thinking through a multi-step solution process	0.304***	0.050*	0.271***	0.031
	Employing conceptual understanding/reasoning	0.268***	0.069***	0.249***	0.070***
	Using reverse thinking/flexible thinking	0.325***	0.084***	0.274***	0.051*
	Making sense of non-routine situations	0.265***	0.021	0.228***	-0.003

Notes: ***, **, * denotes statistical significance at the 1%, 5% and 10% level, respectively.

Individual items

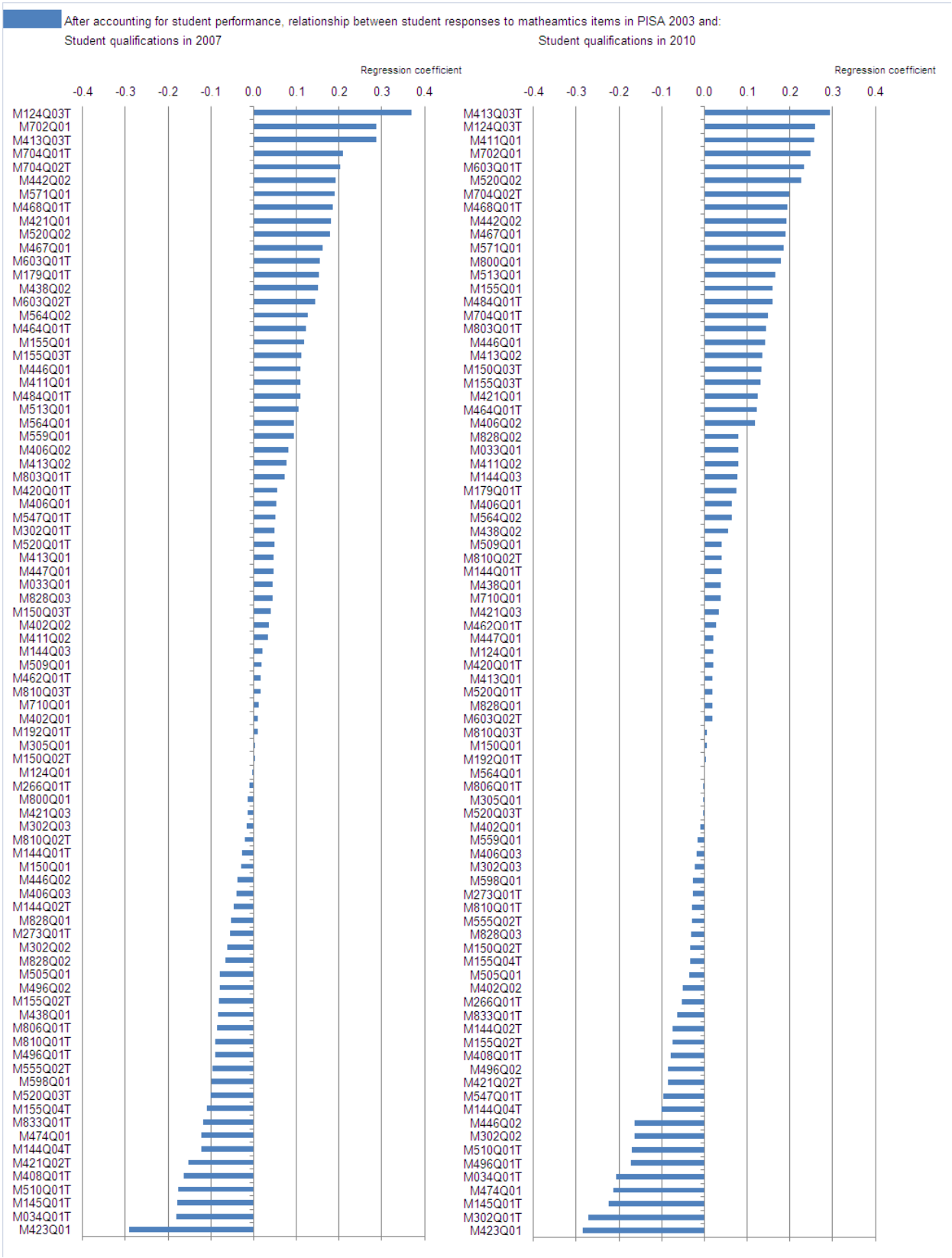
33. Similar analysis is possible for individual items, although in this case sample sizes are generally much smaller as not all students respond to all test items. Relationships between student responses to mathematics items in PISA 2003 and student qualifications measured in 2007 and 20010 are presented in Table A5 in Annex A.[7] The positive coefficient means that a correct response to an item is positively associated with higher student qualifications. The coefficients become smaller in general for all items after accounting for overall performance. However, those items with more positive coefficients than other items in the simple regression model still tend to have more positive coefficients even in the regression model after accounting for overall performance. In fact, over 60% of the variation in the coefficients of the model accounting for overall performance can be explained by the coefficients of the simple regression model.[8]

34. Figure 1 shows the relationship between students' responses to mathematics items in PISA 2003 and their qualifications measured in 2007 and 2010, after accounting for overall performance. In general, the results are consistent between 2007 and 2010. For example, when comparing the top 10 items that most strongly related to qualifications in 2007 and those in 2010, the following seven items are common in both 2007 and 2010: M124Q03T, M413Q03T, M442Q02, M468Q01T, M520Q02, M702Q01 and M704Q02T.

7 . Relationships between student responses to problem-solving items in PISA 2003 and student qualifications measured in 2007 and 20010 are presented in Table A6 in Annex A

8 . Similarly, around 60% of the variation in the coefficients of the model accounting for overall performance and some student background characteristics can be explained by the coefficients of the simple regression model.

Figure 1. Relationship between student responses to mathematics items in PISA 2003 and student qualifications measured in 2007 and 2010



35. Table 6 shows the list of the top ten mathematics items that have the strongest relationships with student qualifications in either 2007 or 2010. The classifications of these items are also presented. Similar to the results based on scales shown in Tables 4, items classified as *statistics* tend to be related to student qualifications. Forty percent of the items classified as *statistics* (four of ten items) are found to be among the mathematics items that have the strongest predictive power as shown in Table 6.

Table 6. Classifications of mathematics with strong relationships with student qualifications either in 2007 or 2010

Unit item code	Original classification				New classification														
	Content Area	Overarching idea	Competency clusters	Item Format	Thinking across multiple representations	Probabilistic reasoning	Statistics	Rates, Ratios, Proportions, %	Measurement	Using a graph, chart, or table	Using multiple graphs	Spatial reasoning	Conceptual understanding vs. procedural knowledge	Argument/explanation	Systematic/sequential/strategic	Studying Examples	Thinking through a multi-step solution process	Employing conceptual understanding/reasoning	Using reverse thinking/flexible thinking	Making sense of non-routine situations
M124Q03T	3	1	1	2	0	0	0	1	0	0	0	0	-1	0	0	0	2	0	0	0
M411Q01	5	2	3	5	0	0	1	0	0	0	0	0	-1	0	0	0	2	0	0	0
M413Q03T	5	2	2	2	0	0	0	1	0	0	0	0	1	1	0	0	0	2	2	2
M421Q01	7	4	3	2	0	0	1	0	0	0	0	0	-1	1	0	0	0	0	2	0
M442Q02	2	2	2	1	0	0	0	0	0	0	0	0	-1	0	1	1	2	0	2	2
M467Q01	6	4	3	3	0	1	0	1	0	1	1	0	-1	0	0	0	1	0	0	0
M468Q01T	5	4	3	5	0	0	1	0	0	0	0	0	-1	0	0	0	2	0	0	0
M520Q02	2	2	3	4	0	0	0	0	0	0	0	0	-1	0	1	0	0	0	0	0
M571Q01	3	4	2	3	0	0	0	0	0	0	0	0	1	0	0	0	0	2	2	0
M603Q01T	5	2	1	4	0	0	0	0	0	0	0	0	-1	0	1	0	0	0	1	2
M702Q01	7	4	1	2	0	0	1	0	0	0	0	0	1	1	0	0	0	2	0	0
M704Q01T	1	1	3	5	0	0	0	0	0	1	1	0	-1	0	0	0	0	0	0	0
M704Q02T	1	1	2	2	0	0	0	0	0	1	1	0	-1	0	1	0	2	0	2	2

Coding information		Coding information		Coding information		Coding information	
Content Area		**Item Format**		**Using a Graph**		**Thinking through a multi-step solution process**	
Algebra	1	Closed Constructed Respo	1	No display	0	**Employing conceptual understanding/reasonin**	
Discrete Mathema	2	Open Constructed Respons	2	Reading a display	1	**Using reverse thinking/flexible thinking**	
Functions	3	Multiple Choice	3	Interpreting a display	2	**Making sense of non-routine situations**	
Geometry	4	Complex Multiple Choice	4	**Using multiple graphs**		Does not contribute to the cognitive load/demand f	0
Number	5	Short Response	5	No graphic organisers	0	Present in an item but does not determine its cogr	1
Probabilistic reas	6	**Competency clusters**		One graphic organiser	1	Contributes to the cognitive load/demand for an ite	2
Statistics	7	Connections	1	Multiple graphic organis	2	**Other than above**	
Overarching idea		Reflection	2	**Conceptual understanding**		Item does not belong to the attribute	0
Change and Rela	1	Reproduction	3	**vs. procedural knowledge**		Item does belongs to the attribute	1
Quantity	2			Procedural knowledge	-1		
Space and Shape	3			Conceptual understand	1		
Uncertainty	4						

36. Two examples of items in the classification of *statistics* are presented in Boxes 1 and 2. One example of items in the classification of *rates, ratios, proportions and percentages* is in Box 3. Examples of items in other classifications are shown in Annex B.

Box 1. Mathematics item: M702Q01

SUPPORT FOR THE PRESIDENT

Question 1: SUPPORT FOR THE PRESIDENT

M702Q01

In Zedland, opinion polls were conducted to find out the level of support for the President in the forthcoming election. Four newspaper publishers did separate nationwide polls. The results for the four newspaper polls are shown below:

Newspaper 1: 36.5% (poll conducted on January 6, with a sample of 500 randomly selected citizens with voting rights)

Newspaper 2: 41.0% (poll conducted on January 20, with a sample of 500 randomly selected citizens with voting rights)

Newspaper 3: 39.0% (poll conducted on January 20, with a sample of 1000 randomly selected citizens with voting rights)

Newspaper 4: 44.5% (poll conducted on January 20, with 1000 readers phoning in to vote).

Which newspaper's result is likely to be the best for predicting the level of support for the President if the election is held on January 25? Give two reasons to support your answer.

……
……

Box 2. Mathematics item: M468Q01T

SCIENCE TESTS

Question 1: SCIENCE TESTS

M468Q01

In Mei Lin's school, her science teacher gives tests that are marked out of 100. Mei Lin has an average of 60 marks on her first four Science tests. On the fifth test she got 80 marks.

What is the average of Mei Lin's marks in Science after all five tests?

Average: .....

Box 3. Mathematics item: M413Q03T

EXCHANGE RATE

Mei-Ling from Singapore was preparing to go to South Africa for 3 months as an exchange student. She needed to change some Singapore dollars (SGD) into South African rand (ZAR).

Question 3: EXCHANGE RATE

M413Q03

During these 3 months the exchange rate had changed from 4.2 to 4.0 ZAR per SGD.

Was it in Mei-Ling's favour that the exchange rate now was 4.0 ZAR instead of 4.2 ZAR, when she changed her South African rand back to Singapore dollars? Give an explanation to support your answer.

……
……
……

37. The relationship between correct responses to items and future qualifications can partly depends on item difficulty and discriminatory power. In general, more difficult items and items with higher discriminatory power are considered to be more highly related to better future outcomes. Figure 2 shows how the mathematics items' levels of difficulty are related to their predictive power. Regression coefficients for individual items after accounting for overall performance are used for measuring items' predictive power. This figure shows that there is almost no relationship between these two: only 3% of the variation in items' predictive power can be explained by the items' level of difficulty.[9] Among items of average difficulty, there are items like M411Q01 with high predictive power, and those like M509Q01 for which correct responses in 2003 are not associated with better student qualifications in 2010.

38. Figure 3 presents how two measures of an item's predictive power are correlated with each other: one is a mathematics item's relationship with student qualifications measured in 2010 (*i.e.* regression coefficient), after accounting for overall performance. The other is a mathematics item's relationship with student qualifications measured in 2010, after accounting for item difficulty and discrimination. These two are highly correlated (R^2 =0.6). In general, items that are strong predictors of student qualifications after accounting for overall performance and student characteristics are still among the best predictors, even after accounting for item IRT parameters. For example, M411Q01 is a strong predictor on both measures.

9 . When regression coefficients for individual items of the simple regression model are used for measuring items' predictive power, 2% of the variation in items' predictive power can be explained by the items' level of difficulty. When regression coefficients for individual items are used for measuring items' predictive power, after accounting for overall performance and some student background characteristics, 4% of the variation in items' predictive power can be explained by the items' level of difficulty.

Figure 2. Correlation between the level of difficulty of a mathematics item and the item's predictive power

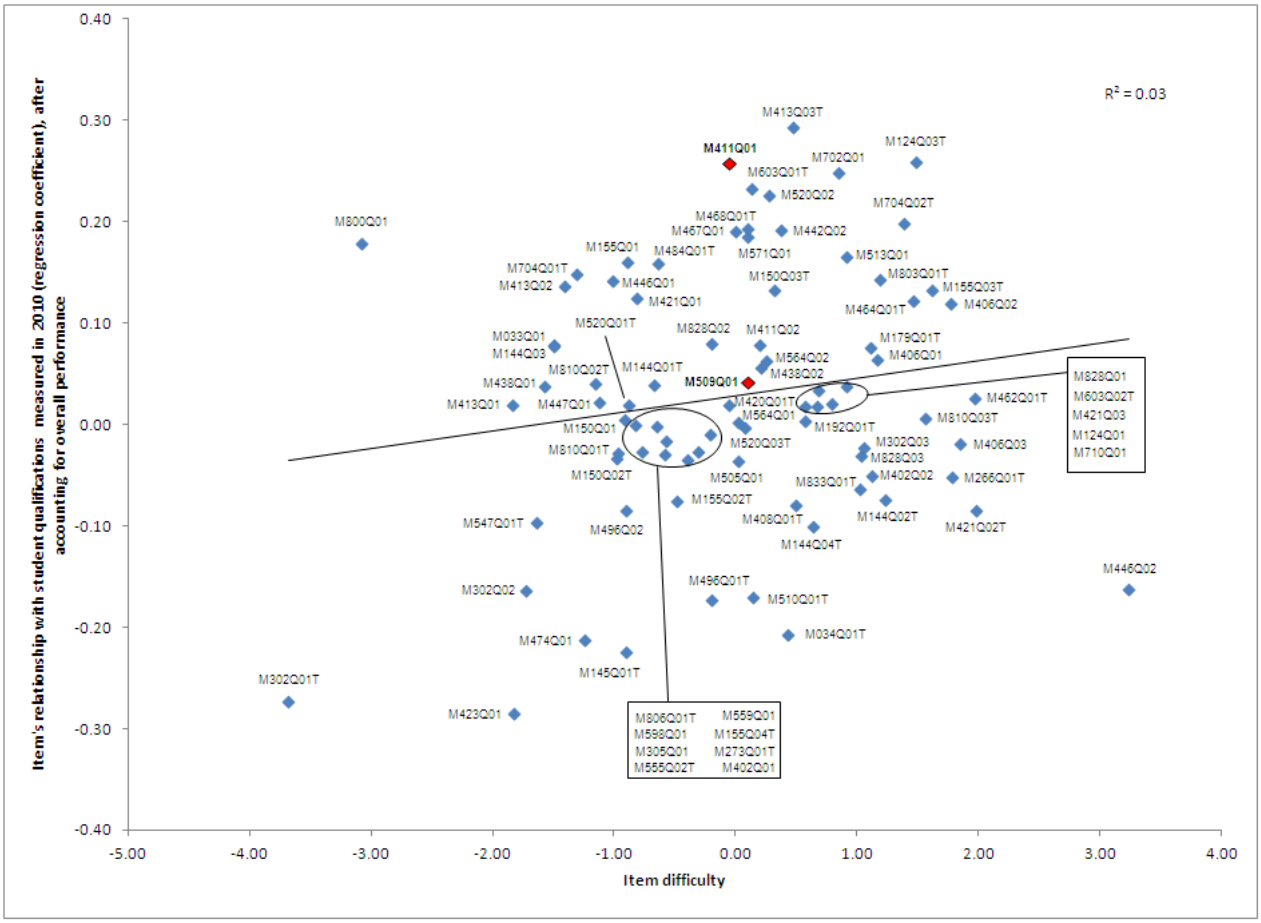
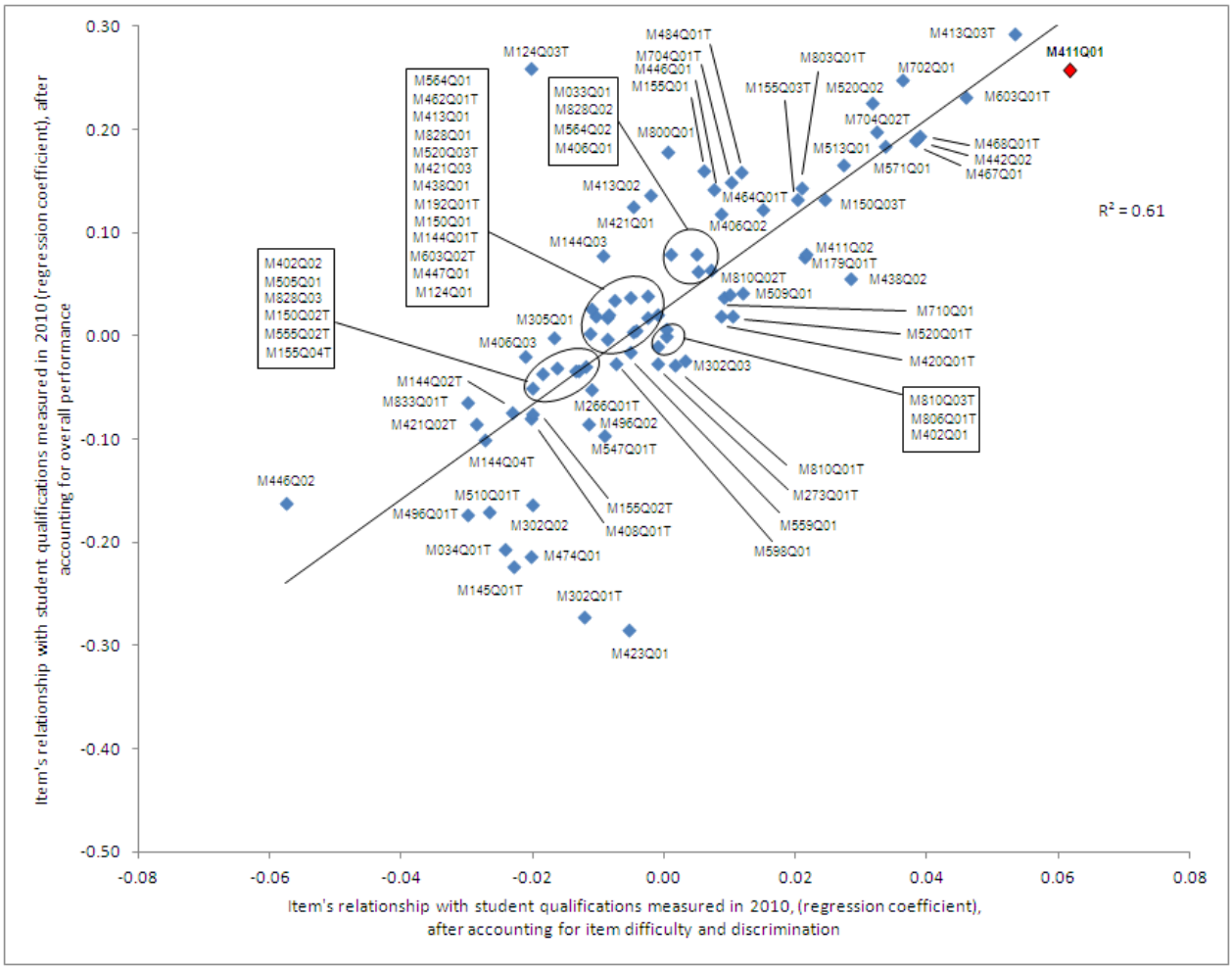


Figure 3. Correlation between two different measures of the predictive power of a mathematics item



39. Similar analysis is possible for problem-solving items in the PISA 2003 study. The results for the problem-solving items are presented in Figures A1 to A3 in Annex A.

Results for item format

40. PISA items have five different formats: short response, multiple choice, complex multiple choice, closed-constructed response, and open-constructed response. It is often argued that simpler formats, like short response or multiple choice, are not able to measure the most advanced skills. Thus, PISA also uses more complex items based on open constructed responses that are then carefully coded. The analysis presented in Table 7 suggests, however, that only positive responses on multiple choice and short-response items are related to student qualifications.[10] This suggests that items that have simpler formats can be at least similarly useful as those of more complex formats.

10. In Table A7 in Annex A, the third set of results is also presented. These are the regression coefficients when the student's gender, socio-economic background and immigrant status were taken into account in addition to overall student performance.

41. Given the small sample size, it is difficult to make generalisations about the nature of these two attribute categories – *statistics* and *rates, ratios, proportions, and percentages* – that make them predictive in terms of item format. Ten items were coded as assessing *statistics*: two multiple-choice items, two complex multiple-choice items, two short-response items, and four open-constructed response items. Ten items were coded as assessing *rates, ratios, proportions, and percentages*: three multiple-choice items, three short-response items, and four open-constructed response items. In sum, of 20 items coded as *statistics* or *rates, ratios, proportions, and percentages*, the largest proportion of items (8 items, or 40%) are open-constructed response items. Five items (25%) are short response, another five items are multiple choice (25%), and the remaining 2 items (10%) are complex multiple choice.

Table 7. Relationship between responses to items of different formats in PISA 2003 and student qualifications measured in 2007 and 2010.

item format	2007		2010	
	Simple regression	After accounting for overall performance (1st plausible value)	Simple regression	After accounting for overall performance (1st plausible value)
	Coef.	Coef.	Coef.	Coef.
		After accounting for mathematics		After accounting for mathematics
Closed constructed response	0.216***	-0.006	0.192***	-0.011
Complex multiple choice	0.228***	-0.004	0.207***	0.004
Multiple choice	0.349***	0.070**	0.295***	0.045
Open constructed response	0.228***	-0.009	0.199***	-0.006
Short response	0.306***	0.045*	0.289***	0.057*
		Accounting for problem-solving performance		Accounting for problem-solving performance
Closed constructed response	0.216***	0.025	0.192***	0.013
Complex multiple choice	0.228***	0.033	0.207***	0.024
Multiple choice	0.349***	0.095***	0.295***	0.060*
Open constructed response	0.228***	0.02	0.199***	0.015
Short response	0.306***	0.081***	0.289***	0.082**

Notes: ***, **, * denotes statistical significance at the 1%, 5% and 10% level, respectively.

V. Conclusions

42. Several countries and economies that have participated in PISA have also established longitudinal studies based on sampled cohorts of 15-year-old students tested in PISA. Longitudinal follow-up studies of students who participated in PISA offer the possibility of exploring how competencies at age 15 are related to educational careers and labour-market outcomes. This report analyses the longitudinal data available from the Australian study (LSAY) that is based on the PISA 2003 cohort. This survey collected information on students' educational careers after the age of 15, which are usually highly associated with future labour market outcomes.

43. For the purpose of this study, several new classifications of PISA mathematics and problem solving items were proposed. Using the student-level micro data, the items were tested to determine whether these new classifications are in fact related to differences in student educational careers. The original PISA assessment framework classifications were also used for this purpose. Unfortunately, it still

not possible to analyse labour-market outcomes for students tested in PISA 2003 as the latest data are available for 2010, and most of these students were not yet looking for full-time employment.

44. Among the newly proposed classifications that can be generalised to other assessment areas besides mathematics, *using reverse thinking/flexible thinking* was found to be significantly related to student qualifications, both in mathematics and problem-solving scales. The problem-solving scale of *employing conceptual understanding/reasoning* is significantly related to student qualifications. Weaker relationships were also found for mathematics items belonging to classes labelled *constructing or presenting an argument or explanation* and problem-solving items belonging to classes labelled *thinking through a multi-step solution process.*

45. From the four original PISA performance subscales representing overarching ideas, only the *space and shape* subscale seems to be unrelated to student qualifications in 2010. In general, however, all four original PISA subscales are positively related to higher student qualifications.

46. It was found that items classified as assessing *statistics* or *rates, ratios, proportions, and percentages* reveal some characteristics that might contribute to their predictiveness. Collectively, these *statistics* items all require students to make sense of the information they are given and apply their knowledge and skills in a way that extends beyond the routine application of a well-defined algorithm. Eight of the ten items involve the concept of average or mean but extend beyond the straight-forward application of the algorithm by eliciting more complex reasoning and thinking. For example, two items require students to reason about the effect of an outlier on the mean; three items require students to determine the mean after a data point has been added or changed. The predictiveness of this set of items appears to lie in the ability of prepared students to be flexible in their thinking and to apply common mathematical concepts to the solution of multi-step, non-routine problems.

47. Ten of the PISA 2003 mathematics items were coded as assessing *rates, ratios, proportions, and percentages*. As with the *statistics* items, all of these items require students to make sense of the information they are given, rather than simply applying a well-defined algorithm to a routine problem. Seven of the items are complex and require multiple steps to arrive at a solution. For example, one open-response item requires students not only to substitute a formula to derive a solution, but to then go through multiple unit conversions before arriving at a solution expressed in the appropriate unit. A number of these items also involve contexts that may not be uniformly familiar to 15-year-old students across a range of countries. Such contexts include the notion of gear ratio, currency exchange rates, population pyramids, and percentage capacity. The predictiveness of this set of items appears to lie in the ability of prepared students to persist through the multiple steps required to solve non-routine problems, and understand and interpret information presented in an unfamiliar format or context.

48. The results of this analysis are promising, although some limitations should be clearly noted. The largest limitation is lack of real measures of future labour-market outcomes of students tested in PISA. For the majority of students, it is still too early to see these outcomes as these students are continuing their education. The outcome measures used in this study are limited to educational outcomes only. With large number of students obtaining the highest qualifications level, it is impossible to see the differences between items in predicting the most desirable outcomes. With future data that measure outcomes in the labour market it will be possible to differentiate later student outcomes and see what items best predict the most desirable life achievements.

REFERENCES

National Research Council (2001), *Adding It Up: Helping Children Learn Mathematics*, National Academy Press, Washington, D.C.

Webb, Norman (2001), *Levels for Determining Depth of Knowledge*, CCSSO TILSA Alignment Study, Version 2.0.

Kolenikov, S. and G. Angeles (2004).,"The Use of Discrete Data in Principal Component Analysis with Applications to Socio-Economic Indices", CPC/MEASURE working paper No. WP-04-85.

OECD (2013), *PISA 2012 Assessment and Analytical Framework*, OECD Publishing.

ANNEX A

Table A1 [1/3]. Mathematics item classifications

Unit item code	Item Name	Original classification: Content Area	Original classification: Overarching idea	Original classification: Competency clusters	Item Format	New classification: Thinking across multiple representations	New classification: Probabilistic reasoning	New classification: Statistics	New classification: Rates, Ratios, Proportions, %	New classification: Measurement	New classification: Using a graph, chart, or table	New classification: Using multiple graphs	New classification: Spatial reasoning	New classification: Conceptual understanding vs. procedural knowledge	New classification: Argumentexplanation	New classification: Systematicsequentialstrategic	New classification: Studying Examples	New classification: Thinking through a multi-step solution process	New classification: Employing conceptual understandingreasoning	New classification: Using reverse thinkingflexible thinking	New classification: Making sense of non-routine situations
M033Q01	P2000 A View Room	4	3	3	3	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	2	2
M034Q01T	P2000 Bricks	4	3	1	1	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0
M124Q01	P2000 Walking	3	1	3	2	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0
M124Q03T	P2000 Walking	3	1	1	2	0	0	0	1	0	0	0	0	-1	0	0	0	2	0	0	0
M144Q01T	P2000 Cube Painting	4	3	3	1	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0
M144Q02T	P2000 Cube Painting	4	3	1	1	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0
M144Q03	P2000 Cube Painting	4	3	1	3	0	0	0	0	0	0	0	1	-1	0	0	0	2	0	0	2
M144Q04T	P2000 Cube Painting	4	3	1	1	0	0	0	0	0	0	0	1	-1	0	1	0	0	0	0	0
M145Q01T	P2000 Cubes	5	3	3	1	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0
M150Q01	P2000 Growing Up	5	1	3	1	1	0	0	0	0	1	0	0	-1	0	0	0	0	0	0	0
M150Q02T	P2000 Growing Up	3	1	3	1	0	0	0	0	0	2	1	0	-1	0	0	0	0	0	0	0
M150Q03T	P2000 Growing Up	3	1	1	2	0	0	0	0	0	1	1	0	-1	0	0	0	0	0	0	0
M155Q01	P2000 Population Pyramids	7	1	1	2	0	0	0	0	0	2	2	0	-1	0	0	0	0	0	0	0
M155Q02T	P2000 Population Pyramids	7	1	1	2	0	0	0	0	0	2	2	0	-1	0	0	0	0	0	0	0
M155Q03T	P2000 Population Pyramids	7	1	2	2	0	0	0	1	0	2	2	0	-1	0	0	0	0	0	0	0
M155Q04T	P2000 Population Pyramids	7	1	1	4	0	0	0	0	0	2	2	0	-1	0	0	0	0	0	0	0
M179Q01T	P2000 Robberies	7	4	1	2	0	0	0	0	0	2	1	0	1	0	0	0	0	0	0	0
M192Q01T	P2000 Containers	3	1	1	4	1	0	0	0	0	2	2	0	-1	0	0	0	2	0	0	0
M266Q01T	P2000 Carpenter	4	3	1	4	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0
M273Q01T	P2000 Pipelines	4	3	1	4	0	0	0	0	0	0	0	0	-1	0	0	0	2	0	0	0
M302Q01T	Car Drive	3	1	3	1	0	0	0	0	0	1	1	0	-1	0	0	0	0	0	0	0
M302Q02	Car Drive	3	1	1	1	0	0	0	0	0	2	1	0	1	0	0	0	0	1	0	0
M302Q03	Car Drive	3	1	2	2	0	0	0	0	0	2	1	0	1	0	0	0	0	2	0	0
M305Q01	Map	4	3	1	3	0	0	0	0	1	0	0	0	-1	0	0	0	0	0	0	0
M402Q01	Internet Relay Chat	5	1	1	5	0	0	0	0	1	0	0	0	-1	0	0	0	0	0	0	0
M402Q02	Internet Relay Chat	5	1	2	5	0	0	0	0	1	0	0	0	-1	0	1	0	2	0	1	0
M406Q01	Running Tracks	4	3	1	2	0	0	0	0	1	0	0	0	-1	0	0	0	1	1	0	0
M406Q02	Running Tracks	4	3	1	2	0	0	0	0	1	0	0	0	-1	0	0	0	2	2	0	0
M406Q03	Running Tracks	4	3	2	2	0	0	0	0	1	0	0	0	-1	0	0	0	0	2	0	0
M408Q01T	Lotteries	6	4	1	4	0	1	0	0	0	0	0	0	-1	0	0	0	0	0	0	0
M411Q01	Diving	5	2	3	5	0	0	1	0	0	0	0	0	-1	0	0	0	2	0	0	0
M411Q02	Diving	7	4	1	3	0	0	1	0	0	0	0	0	-1	0	0	0	0	2	0	0
M413Q01	Exchange Rate	5	2	3	5	0	0	0	1	0	0	0	0	-1	0	0	0	0	0	0	0
M413Q02	Exchange Rate	5	2	3	5	0	0	0	1	0	0	0	0	-1	0	0	0	0	0	1	2
M413Q03T	Exchange Rate	5	2	2	2	0	0	0	1	0	0	0	0	-1	0	0	0	0	2	2	2
M420Q01T	Transport	7	4	2	4	0	0	1	0	0	2	1	0	-1	0	0	0	0	2	1	0
M421Q01	Height	7	4	3	2	0	0	1	0	0	0	0	0	-1	0	1	0	0	0	2	0
M421Q02T	Height	7	4	2	4	0	0	1	0	0	0	0	0	-1	0	1	0	0	2	2	0
M421Q03	Height	7	4	2	3	0	0	1	0	0	0	0	0	-1	0	0	0	1	1	2	0
M423Q01	Tossing Coins	6	4	3	3	0	1	0	0	0	0	0	0	-1	0	0	0	0	2	0	0
M438Q01	Exports	7	4	3	1	0	0	0	0	0	2	0	0	-1	0	0	0	0	0	0	0
M438Q02	Exports	7	4	1	3	1	0	0	0	0	2	0	0	-1	0	0	0	2	0	0	0

Table A1 [2/3]. Mathematics item classifications

Unit item code	Item Name	Original classification: Content Area	Original classification: Overarching idea	Original classification: Competency clusters	Item Format	New classification: Thinking across multiple representations	New classification: Probabilistic reasoning	New classification: Statistics	New classification: Rates, Ratios, Proportions, %	New classification: Measurement	New classification: Using a graph, chart, or table	New classification: Using multiple graphs	New classification: Spatial reasoning	New classification: Conceptual understanding vs. procedural knowledge	New classification: Argumentexplanation	New classification: Systematicsequentialstrategic	New classification: Studying Examples	New classification: Thinking through a multi-step solution process	New classification: Employing conceptual understandingfreasoning	New classification: Using reverse thinkingflexible thinking	New classification: Making sense of non-routine situations
M442Q02	Braille	2	2	2	1	0	0	0	0	0	0	0	0	-1	0	1	1	2	0	2	2
M446Q01	Thermometer Cricket	5	1	3	5	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0
M446Q02	Thermometer Cricket	1	1	2	2	1	0	0	0	0	0	0	0	-1	0	0	0	0	0	2	0
M447Q01	Tile Arrangement	4	3	3	3	0	0	0	0	0	0	0	0	-1	0	0	0	0	2	2	2
M462Q01T	Third Side	4	3	2	2	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
M464Q01T	The Fence	4	3	1	5	0	0	0	0	1	0	0	0	-1	0	0	0	1	0	1	0
M467Q01	Coloured Candles	6	4	3	3	0	0	0	1	0	1	1	0	-1	0	0	0	1	0	0	0
M468Q01T	Science Tests	5	4	3	5	0	0	1	0	0	0	0	0	-1	0	0	0	2	0	0	0
M474Q01	Running Time	5	2	3	1	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	1	0
M484Q01T	Bookshelves	5	2	1	5	0	0	0	0	0	0	0	0	-1	0	0	0	2	0	0	0
M496Q01T	Cash Withdrawal	5	2	1	4	0	0	0	0	0	0	0	0	-1	0	0	0	2	0	0	0
M496Q02	Cash Withdrawal	5	2	1	5	0	0	0	0	0	0	0	0	-1	0	0	0	2	0	0	0
M505Q01	Litter	7	4	2	2	0	0	1	0	0	2	1	0	-1	1	0	0	0	0	1	2
M509Q01	Earthquake	6	4	2	3	0	1	0	0	0	0	0	0	-1	0	0	0	0	0	0	1
M510Q01T	Choices	2	2	1	5	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0
M513Q01	Test Scores	7	4	1	2	0	0	1	0	0	2	1	0	-1	1	0	0	0	2	0	0
M520Q01T	Skateboard	5	2	3	5	0	0	0	0	0	1	1	0	-1	0	0	0	2	0	0	2
M520Q02	Skateboard	2	2	3	4	0	0	0	0	0	0	0	0	-1	0	0	1	0	0	0	0
M520Q03T	Skateboard	5	2	1	5	0	0	0	0	0	1	1	0	-1	0	0	1	2	0	2	0
M547Q01T	Staircase	5	3	3	5	0	0	0	0	1	0	0	0	-1	0	0	0	0	0	0	0
M555Q02T	Number Cubes	4	3	1	4	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0
M559Q01	Telephone Rates	5	2	2	3	0	0	0	0	0	1	1	0	-1	0	0	0	0	0	0	2
M564Q01	Chair Lift	5	2	3	3	0	0	0	1	0	0	0	0	-1	0	0	0	2	0	0	0
M564Q02	Chair Lift	2	4	2	3	0	0	0	1	0	0	0	0	-1	0	0	0	2	0	0	0
M571Q01	Stop The Car	3	4	2	3	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0
M598Q01	Making A Booklet	4	3	2	1	0	0	0	0	0	0	0	1	-1	0	0	0	1	0	1	2
M603Q01T	Number Check	5	2	1	4	0	0	0	0	0	0	0	0	-1	0	0	1	0	0	1	0
M603Q02T	Number Check	5	2	1	5	0	0	0	0	0	0	0	0	-1	0	0	0	2	0	0	0
M702Q01	Support For President	7	4	1	2	0	0	1	0	0	0	0	0	1	0	0	0	0	2	0	0
M704Q01T	The Best Car	1	1	3	5	0	0	0	0	0	1	1	0	-1	0	0	0	0	0	0	0
M704Q02T	The Best Car	1	1	2	2	0	0	0	0	0	1	1	0	-1	0	0	0	2	0	2	0
M710Q01	Forecast of Rain	6	4	1	3	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0
M800Q01	Computer Game	5	2	3	3	0	0	0	0	0	1	1	0	-1	0	0	0	0	0	0	0
M803Q01T	Labels	7	4	1	5	0	0	0	1	0	0	0	0	-1	0	0	0	2	0	0	0
M806Q01T	Step Pattern	2	2	3	5	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0
M810Q01T	Bicycles	5	2	1	5	0	0	0	0	0	1	1	0	-1	0	0	0	0	0	0	0
M810Q02T	Bicycles	5	2	1	5	0	0	0	0	0	1	1	0	-1	0	0	0	0	0	0	1
M810Q03T	Bicycles	5	1	2	2	0	0	0	1	0	0	0	0	-1	0	0	0	2	0	0	2
M828Q01	Carbon Dioxide	7	1	3	2	0	0	0	0	0	1	2	0	-1	0	0	0	0	0	0	0
M828Q02	Carbon Dioxide	7	4	1	5	0	0	0	0	0	2	2	0	-1	0	0	0	1	0	0	0
M828Q03	Carbon Dioxide	5	2	1	5	0	0	0	0	0	2	1	0	-1	0	0	0	1	0	0	0
M833Q01T	Seeing the tower	4	3	1	4	0	0	0	0	0	2	2	1	-1	0	0	0	0	0	0	1

Table A1 [3/3]. Mathematics item classifications

Coding information			
Content Area			
Algebra	1		
Discrete Mathematics	2		
Functions	3		
Geometry	4		
Number	5		
Probabilistic reasoning	6		
Statistics	7		
Overarching idea		**Year**	**Section**
Change and Relationships	1	2003	Math
Quantity	2	2003	Math
Space and Shape	3	2000 and 2003	Math
Uncertainty	4	2003	Math
Item Format			
Closed Constructed Response	1		
Open Constructed Response	2		
Multiple Choice	3		
Complex Multiple Choice	4		
Short Response	5		
Competency clusters			
Connections	1		
Reflection	2		
Reproduction	3		
Using a Graph			
No display	0		
Reading a display	1		
Interpreting a display	2		
Using multiple graphs			
No graphic organisers	0		
One graphic organiser	1		
Multiple graphic organisers	2		
Conceptual understanding vs. procedural knowledge			
Procedural knowledge	-1		
Conceptual understanding	1		
Thinking through a multi-step solution process **Employing conceptual understanding/reasoning** **Using reverse thinking/flexible thinking** **Making sense of non-routine situations**			
Does not contribute to the cognitive load/demand for an item			0
Present in an item but does not determine its cognitive load/demand			1
Contributes to the cognitive load/demand for an item			2
Other than above			
Item does not belong to the attribute	0		
Item does belongs to the attribute	1		

Table A2 [1/2]. Problem-solving item classifications

Unit item code	Item Name	Original classification: Overarching idea	Item Format	New classification: Thinking across multiple representations	Probabilistic reasoning	Statistics	Rates, Ratios, Proportions, %	Measurement	Using a graph, chart, or table	Using multiple graphs	Spatial reasoning	Conceptual understanding vs. procedural knowledge	Argumentexplanation	Systematic/sequential/strategic	Studying Examples	Thinking through a multi-step solution process	Employing conceptual understanding/reasoning	Using reverse thinking/flexible thinking	Making sense of non-routine situations
X402Q01T	Library System	6	1	0	0	0	0	0	0	0	0	-1	0	0	0	2	0	0	0
X402Q02T	Library System	6	2	0	0	0	0	0	0	0	0	-1	0	1	1	2	0	0	2
X412Q01	Design by Numbers	6	3	0	0	0	0	0	2	1	0	-1	0	0	1	0	0	1	1
X412Q02	Design by Numbers	6	3	0	0	0	0	0	2	1	0	-1	0	1	1	2	0	1	2
X412Q03	Design by Numbers	6	2	0	0	0	0	0	2	1	0	-1	0	1	1	2	0	0	2
X414Q01	Course Design	6	2	0	0	0	0	0	0	1	0	-1	0	0	1	2	0	0	2
X415Q01T	Transit System	5	2	0	0	0	0	0	1	1	1	-1	0	0	0	2	0	0	2
X417Q01	Children's Camp	6	2	0	0	0	0	0	1	2	0	-1	0	1	0	2	0	0	2
X423Q01T	Freezer	7	3	0	0	0	0	0	0	0	0	-1	0	1	0	0	1	1	1
X423Q02T	Freezer	7	3	0	0	0	0	0	2	1	0	-1	0	1	0	0	2	0	1
X430Q01	Energy Needs	5	1	0	0	0	0	0	1	2	0	-1	0	0	0	2	0	0	2
X430Q02	Energy Needs	5	2	0	0	0	0	0	1	2	0	-1	0	1	0	2	0	1	2
X601Q01T	Cinema Outing	5	3	0	0	0	0	0	1	1	0	-1	0	0	0	2	0	1	2
X601Q02	Cinema Outing	5	3	0	0	0	0	0	1	1	0	-1	0	0	0	2	0	1	1
X602Q01	Holiday	5	1	0	0	0	0	0	1	2	0	-1	0	0	0	2	0	0	0
X602Q02	Holiday	5	2	0	0	0	0	0	1	2	0	-1	0	0	0	2	0	0	2
X603Q01	Irrigation	7	2	0	0	0	0	0	1	2	0	-1	0	1	0	2	0	0	1
X603Q02T	Irrigation	7	3	0	0	0	0	0	2	2	0	-1	0	1	0	0	0	2	1
X603Q03	Irrigation	7	2	0	0	0	0	0	2	2	0	-1	0	1	0	2	0	0	2

Table A2 [2/2]. Problem-solving item classifications

Coding information			
Overarching idea		**Year**	**Section**
Decision Making	5	2003	PS
System Analysis and Design	6	2003	PS
Troubleshooting	7	2003	PS
Item Format			
Closed Constructed Response	1		
Open Constructed Response	2		
Multiple Choice	3		
Complex Multiple Choice	4		
Short Response	5		
Competency clusters			
Connections	1		
Reflection	2		
Reproduction	3		
Using a Graph			
No display	0		
Reading a display	1		
Interpreting a display	2		
Using multiple graphs			
No graphic organisers	0		
One graphic organiser	1		
Multiple graphic organisers	2		
Conceptual understanding vs. procedural knowledge			
Procedural knowledge	-1		
Conceptual understanding	1		
Thinking through a multi-step solution process			
Employing conceptual understanding/reasoning			
Using reverse thinking/flexible thinking			
Making sense of non-routine situations			
Does not contribute to the cognitive load/demand for an item			0
Present in an item but does not determine its cognitive load/demand			1
Contributes to the cognitive load/demand for an item			2
Other than above			
Item does not belong to the attribute	0		
Item does belongs to the attribute	1		

Table A3. Relationship between student responses to various mathematic scales in PISA 2003 and student qualifications measured in 2007 and 2010: three regression models

Types of classifications	Items classification (scale)	2007				2010			
		Simple regression	After accounting for overall performance (1st plausible value in mathematics)	After accounting for overall performance and gender, socio-economic and immigrant background		Simple regression	Adjusted for overall performance (1st plausible value in mathematics)	Adjusted for overall performance and gender, socio-economic and immigrant background	
		Coef.	Coef.	Coef.		Coef.	Coef.	Coef.	
Newly developed classifications	Thinking across multiple representations	0.131***	0.013	0.028		0.102***	-0.001	0.012	
	Probabilistic reasoning	0.183***	0.005	0.014		0.182***	0.019	0.022	
	Statistics	0.283***	0.068***	0.066***		0.267***	0.077**	0.066**	
	Rates, ratios, proportions, and percentages	0.220***	0.059**	0.062***		0.194***	0.047*	0.045*	
	Measurement	0.180***	0.014	0.020		0.169***	0.021	0.019	
	Using graphs, charts, tables, pictures, and diagrams	0.309***	0.003	0.036		0.272***	0.002	0.012	
	Multiple graphs, charts, tables, pictures, and/or diagrams	0.309***	0.003	0.036		0.272***	0.002	0.012	
	Spatial reasoning	0.193***	-0.016	-0.005		0.165***	-0.023	-0.022	
	Conceptual understanding vs. procedural knowledge	0.272***	0.035	0.054**		0.243***	0.033	0.045*	
	Constructing or presenting an argument or explanation	0.238***	0.047*	0.051**		0.206***	0.038	0.035	
	Systematic, sequential, or strategic reasoning	0.245***	-0.001	0.025		0.209***	-0.014	-0.004	
	Applying information/learning through studying examples	0.126***	-0.003	0.006		0.115***	0.001	0.000	
	Thinking through a multi-step solution process	0.304***	0.022	0.041		0.271***	0.021	0.013	
	Employing conceptual understanding/reasoning	0.268***	0.030	0.046*		0.249***	0.040	0.047*	
	Using reverse thinking/flexible thinking	0.325***	0.065**	0.082***		0.274***	0.041	0.049*	
	Making Sense of Non-Routine Situations	0.265***	0.004	0.021		0.228***	-0.005	-0.008	
Overarching idea	Space and shape	0.451***	0.072*	0.099***		0.409***	0.072	0.068	
	Change and relationship	0.503***	0.265***	0.250***		0.460***	0.238***	0.211***	
	Uncertainty	0.514***	0.297***	0.322***		0.462***	0.242***	0.243***	
	Quantity	0.490***	0.211***	0.193***		0.445***	0.185***	0.144***	
Content area	Algebra	0.125***	0.036	0.037*		0.078***	0.006	0.006	
	Discrete mathematics	0.233***	0.035	0.033		0.209***	0.040	0.035	
	Functions	0.224***	0.022	0.036*		0.196***	0.010	0.027	
	Geometry	0.226***	0.000	0.005		0.208***	0.011	0.005	
	Number	0.309***	0.059*	0.076**		0.282***	0.049	0.057*	
	Probabilistic reasoning	0.183***	0.005	0.014		0.182***	0.019	0.022	
	Statistics	0.267***	0.025	0.033		0.244***	0.031	0.026	
Competencies	Connections	0.312***	0.029	0.042*		0.285***	0.038	0.036	
	Reflection	0.236***	0.034	0.045*		0.186***	0.011	0.014	
	Reproduction	0.354***	0.066***	0.076***		0.322***	0.069**	0.075**	

Notes: ***, **, * denotes statistical significance at the 1%, 5% and 10% level, respectively.

Table A4. Relationship between student responses to various problem-solving scales in PISA 2003 and student qualifications measured in 2007 and 2010: three regression models

Types of classifications	Items classification (scale)	2007			2010		
		Simple regression	After accounting for overall performance (1st plausible value in problem solving)	After accounting for overall performance and gender, socio-economic and immigrant background	Simple regression	After accounting for overall performance (1st plausible value in problem solving)	After accounting for overall performance and gender, socio-economic and immigrant background
		Coef.	Coef.	Coef.	Coef.	Coef.	Coef.
Problem types	Decision making	0.119***	-0.034	-0.021	0.116***	-0.030	-0.038
	System analysis and design	0.102***	-0.037	-0.019	0.090***	-0.040*	-0.033
	Troubleshooting	0.112***	-0.041	-0.019	0.080***	-0.065**	-0.051*
Newly developed classifications	Thinking through a multi-step solution process	0.304***	0.050*	0.067**	0.271***	0.031	0.020
	Employing conceptual understanding/reasoning	0.268***	0.069***	0.081***	0.249***	0.070***	0.075***
	Using reverse thinking/flexible thinking	0.325***	0.084***	0.101***	0.274***	0.051*	0.058*
	Making sense of non-routine situations	0.265***	0.021	0.038	0.228***	-0.003	-0.007

Notes: *** ** * denotes statistical significance at the 1%, 5% and 10% level, respectively.

Table A5 [1/2]. Relationships between student responses to mathematics items in PISA 2003 and student qualifications measured in 2007 and 2010

Unit item code	Item Name	2007						2010					
		Simple regression			After accounting for overall performance		After accounting for overall performance, gender, SES and immigrant background	Simple regression			After accounting for overall performance		After accounting for overall performance, gender, SES and immigrant background
		Expl var	Coef.	SE	Coef.	SE	Coef.	SE	Expl var	Coef.	SE	Coef.	SE	Coef.	SE
M033Q01	P2000 A View Room	0.026	0.470	0.093	0.045	0.089	0.014	0.078	0.027	0.468	0.115	0.079	0.116	0.009	0.103
M034Q01T	P2000 Bricks	0.024	0.323	0.066	-0.182	0.067	-0.136	0.060	0.019	0.287	0.076	-0.207	0.082	-0.137	0.076
M124Q01	P2000 Walking	0.049	0.463	0.069	0.000	0.068	0.007	0.067	0.042	0.434	0.080	0.020	0.078	0.003	0.078
M124Q03T	P2000 Walking	0.059	0.495	0.073	0.370	0.086	0.282	0.084	0.041	0.422	0.093	0.259	0.087	0.167	0.082
M144Q01T	P2000 Cube Painting	0.033	0.424	0.081	-0.028	0.078	0.005	0.072	0.033	0.434	0.097	0.039	0.096	0.027	0.091
M144Q02T	P2000 Cube Painting	0.028	0.371	0.068	-0.046	0.066	-0.034	0.061	0.021	0.325	0.080	-0.074	0.079	-0.098	0.075
M144Q03	P2000 Cube Painting	0.040	0.646	0.110	0.020	0.101	-0.016	0.097	0.038	0.633	0.127	0.078	0.134	-0.004	0.130
M144Q04T	P2000 Cube Painting	0.020	0.300	0.080	-0.122	0.073	-0.099	0.069	0.017	0.278	0.085	-0.100	0.081	-0.104	0.078
M145Q01T	P2000 Cubes	0.026	0.416	0.086	-0.179	0.085	-0.109	0.072	0.020	0.355	0.094	-0.224	0.098	-0.178	0.092
M150Q01	P2000 Growing Up	0.035	0.449	0.107	-0.030	0.095	0.010	0.103	0.036	0.448	0.118	0.005	0.110	0.020	0.123
M150Q02T	P2000 Growing Up	0.029	0.307	0.056	0.002	0.048	0.028	0.047	0.016	0.223	0.060	-0.034	0.060	-0.006	0.060
M150Q03T	P2000 Growing Up	0.049	0.494	0.073	0.040	0.068	-0.041	0.069	0.063	0.556	0.084	0.133	0.091	0.063	0.093
M155Q01	P2000 Population Pyramids	0.034	0.491	0.087	0.117	0.085	0.111	0.080	0.033	0.477	0.131	0.160	0.124	0.189	0.111
M155Q02T	P2000 Population Pyramids	0.016	0.190	0.049	-0.081	0.047	-0.075	0.046	0.011	0.154	0.061	-0.075	0.061	-0.099	0.063
M155Q03T	P2000 Population Pyramids	0.083	0.409	0.043	0.111	0.050	0.072	0.047	0.076	0.388	0.045	0.132	0.052	0.068	0.052
M155Q04T	P2000 Population Pyramids	0.013	0.262	0.076	-0.108	0.071	-0.074	0.062	0.017	0.296	0.098	-0.034	0.098	-0.047	0.084
M179Q01T	P2000 Robberies	0.108	0.451	0.042	0.153	0.050	0.134	0.047	0.070	0.353	0.050	0.076	0.058	0.090	0.055
M192Q01T	P2000 Containers	0.044	0.437	0.063	0.009	0.067	0.066	0.064	0.041	0.419	0.077	0.003	0.086	0.063	0.081
M266Q01T	P2000 Carpenter	0.034	0.454	0.072	-0.009	0.074	0.002	0.074	0.035	0.476	0.083	-0.052	0.086	-0.091	0.086
M273Q01T	P2000 Pipelines	0.018	0.297	0.078	-0.056	0.093	-0.067	0.089	0.028	0.365	0.091	-0.027	0.091	-0.063	0.090
M302Q01T	Car Drive	0.001	0.315	0.327	0.049	0.302	-0.075	0.310	0.000	0.108	0.380	-0.273	0.177	-0.459	0.182
M302Q02	Car Drive	0.004	0.186	0.111	-0.061	0.107	0.005	0.108	0.006	0.222	0.122	-0.163	0.103	-0.105	0.106
M302Q03	Car Drive	0.051	0.480	0.073	-0.016	0.076	0.070	0.075	0.054	0.492	0.081	-0.024	0.088	0.023	0.082
M305Q01	Map	0.016	0.299	0.086	0.003	0.087	-0.012	0.087	0.011	0.258	0.087	-0.001	0.084	0.041	0.078
M402Q01	Internet Relay Chat	0.039	0.427	0.069	0.009	0.061	0.042	0.061	0.031	0.379	0.078	-0.009	0.079	0.000	0.075
M402Q02	Internet Relay Chat	0.044	0.442	0.065	0.035	0.065	0.054	0.060	0.025	0.332	0.086	-0.050	0.077	-0.030	0.071
M406Q01	Running Tracks	0.082	0.609	0.072	0.053	0.066	0.053	0.063	0.071	0.571	0.088	0.064	0.081	0.028	0.076
M406Q02	Running Tracks	0.095	0.651	0.067	0.082	0.076	0.133	0.080	0.080	0.607	0.080	0.119	0.078	0.179	0.078
M406Q03	Running Tracks	0.044	0.447	0.079	-0.039	0.074	-0.035	0.073	0.036	0.406	0.103	-0.019	0.094	-0.010	0.093
M408Q01T	Lotteries	0.010	0.218	0.067	-0.162	0.060	-0.142	0.057	0.017	0.272	0.081	-0.080	0.075	-0.089	0.071
M411Q01	Diving	0.086	0.628	0.061	0.110	0.065	0.080	0.062	0.104	0.687	0.073	0.257	0.080	0.228	0.082
M411Q02	Diving	0.051	0.477	0.067	0.033	0.064	0.001	0.061	0.055	0.500	0.080	0.079	0.081	0.041	0.076
M413Q01	Exchange Rate	0.041	0.619	0.080	0.047	0.083	0.062	0.079	0.029	0.535	0.116	0.020	0.116	0.038	0.113
M413Q02	Exchange Rate	0.059	0.723	0.090	0.077	0.094	0.117	0.091	0.045	0.650	0.119	0.136	0.126	0.231	0.122
M413Q03T	Exchange Rate	0.118	0.725	0.060	0.286	0.070	0.224	0.072	0.097	0.656	0.070	0.293	0.076	0.224	0.075
M420Q01T	Transport	0.050	0.521	0.070	0.056	0.068	0.120	0.063	0.043	0.469	0.085	0.020	0.085	0.106	0.082
M421Q01	Height	0.068	0.705	0.087	0.181	0.089	0.102	0.080	0.049	0.582	0.103	0.125	0.114	0.053	0.109
M421Q02T	Height	0.014	0.273	0.080	-0.153	0.070	-0.054	0.064	0.018	0.319	0.086	-0.085	0.078	-0.012	0.078
M421Q03	Height	0.038	0.428	0.070	-0.015	0.072	0.036	0.068	0.029	0.374	0.083	0.034	0.081	0.077	0.078
M423Q01	Tossing Coins	0.002	0.139	0.121	-0.292	0.107	-0.173	0.094	0.004	0.169	0.138	-0.285	0.128	-0.207	0.112
M438Q01	Exports	0.004	0.213	0.168	-0.083	0.194	-0.088	0.187	0.007	0.284	0.140	0.038	0.132	0.050	0.124
M438Q02	Exports	0.105	0.682	0.065	0.150	0.067	0.190	0.065	0.074	0.579	0.075	0.056	0.076	0.094	0.073

Table A5 [2/2]. Relationships between student responses to mathematics items in PISA 2003 and student qualifications measured in 2007 and 2010

Unit item code	Item Name	2007						2010					
		Simple regression			After accounting for overall performance		After accounting for overall performance, gender, SES and immigrant background	Simple regression			After accounting for overall performance		After accounting for overall performance, gender, SES and immigrant background
		Expl var	Coef.	SE	Coef.	SE	Coef.	SE	Expl var	Coef.	SE	Coef.	SE	Coef.	SE
M442Q02	Braille	0.093	0.651	0.063	0.192	0.075	0.135	0.068	0.090	0.636	0.079	0.191	0.092	0.133	0.079
M446Q01	Thermometer Cricket	0.042	0.556	0.087	0.111	0.090	0.160	0.072	0.047	0.567	0.109	0.142	0.108	0.196	0.090
M446Q02	Thermometer Cricket	0.025	0.477	0.088	-0.037	0.096	-0.041	0.091	0.011	0.324	0.103	-0.163	0.105	-0.164	0.100
M447Q01	Tile Arrangement	0.044	0.527	0.083	0.047	0.080	0.040	0.081	0.038	0.481	0.101	0.021	0.105	0.041	0.100
M462Q01T	Third Side	0.049	0.364	0.041	0.017	0.046	0.028	0.047	0.041	0.351	0.055	0.026	0.059	0.038	0.058
M464Q01T	The Fence	0.087	0.644	0.061	0.123	0.067	0.122	0.063	0.078	0.605	0.072	0.122	0.076	0.105	0.073
M467Q01	Coloured Candles	0.083	0.637	0.069	0.162	0.082	0.128	0.076	0.080	0.621	0.086	0.190	0.097	0.158	0.094
M468Q01T	Science Tests	0.071	0.579	0.068	0.186	0.070	0.153	0.067	0.073	0.578	0.088	0.193	0.092	0.145	0.079
M474Q01	Running Time	0.002	0.127	0.076	-0.122	0.078	-0.078	0.078	0.001	0.068	0.087	-0.213	0.085	-0.166	0.083
M484Q01T	Bookshelves	0.066	0.609	0.066	0.109	0.074	0.148	0.069	0.057	0.559	0.086	0.159	0.089	0.148	0.081
M496Q01T	Cash Withdrawal	0.027	0.361	0.070	-0.090	0.072	-0.057	0.062	0.015	0.270	0.084	-0.173	0.082	-0.112	0.078
M496Q02	Cash Withdrawal	0.011	0.240	0.070	-0.079	0.074	-0.092	0.061	0.014	0.260	0.084	-0.085	0.102	-0.106	0.081
M505Q01	Litter	0.015	0.280	0.089	-0.078	0.082	-0.054	0.078	0.013	0.265	0.092	-0.037	0.077	-0.057	0.074
M509Q01	Earthquake	0.041	0.437	0.071	0.017	0.075	-0.020	0.066	0.042	0.444	0.082	0.041	0.086	0.015	0.075
M510Q01T	Choices	0.008	0.188	0.072	-0.176	0.066	-0.204	0.065	0.005	0.147	0.079	-0.171	0.072	-0.207	0.073
M513Q01	Test Scores	0.045	0.451	0.076	0.106	0.071	0.083	0.070	0.065	0.539	0.084	0.165	0.083	0.137	0.083
M520Q01T	Skateboard	0.054	0.349	0.042	0.048	0.047	0.040	0.044	0.045	0.298	0.050	0.019	0.052	0.036	0.050
M520Q02	Skateboard	0.091	0.643	0.066	0.178	0.085	0.172	0.072	0.085	0.614	0.073	0.226	0.102	0.191	0.089
M520Q03T	Skateboard	0.013	0.244	0.067	-0.100	0.062	-0.061	0.057	0.025	0.332	0.075	-0.003	0.069	0.042	0.066
M547Q01T	Staircase	0.032	0.604	0.090	0.051	0.087	0.052	0.088	0.017	0.435	0.136	-0.097	0.124	-0.071	0.127
M555Q02T	Number Cubes	0.035	0.438	0.079	-0.097	0.078	-0.056	0.072	0.037	0.437	0.084	-0.029	0.084	0.018	0.085
M559Q01	Telephone Rates	0.064	0.575	0.072	0.094	0.073	0.116	0.067	0.030	0.397	0.095	-0.016	0.094	0.031	0.092
M564Q01	Chair Lift	0.043	0.440	0.071	0.095	0.074	0.091	0.071	0.022	0.313	0.089	0.002	0.078	-0.009	0.080
M564Q02	Chair Lift	0.053	0.486	0.067	0.127	0.067	0.151	0.059	0.036	0.397	0.075	0.063	0.079	0.096	0.072
M571Q01	Stop The Car	0.093	0.663	0.072	0.190	0.080	0.212	0.074	0.076	0.591	0.084	0.184	0.091	0.243	0.086
M598Q01	Making A Booklet	0.015	0.300	0.073	-0.098	0.067	-0.148	0.066	0.017	0.314	0.084	-0.027	0.083	-0.067	0.080
M603Q01T	Number Check	0.059	0.519	0.072	0.155	0.070	0.102	0.068	0.080	0.604	0.075	0.232	0.075	0.172	0.075
M603Q02T	Number Check	0.048	0.440	0.072	0.145	0.075	0.102	0.073	0.029	0.349	0.103	0.017	0.104	-0.002	0.093
M702Q01	Support For President	0.099	0.661	0.070	0.288	0.075	0.251	0.069	0.083	0.593	0.078	0.248	0.083	0.218	0.077
M704Q01T	The Best Car	0.055	0.748	0.095	0.209	0.089	0.120	0.086	0.043	0.639	0.120	0.149	0.120	0.097	0.113
M704Q02T	The Best Car	0.116	0.731	0.063	0.202	0.076	0.218	0.078	0.093	0.652	0.074	0.198	0.088	0.202	0.096
M710Q01	Forecast of Rain	0.052	0.489	0.066	0.012	0.062	0.022	0.057	0.049	0.482	0.075	0.038	0.076	0.050	0.070
M800Q01	Computer Game	0.006	0.276	0.112	-0.014	0.090	-0.026	0.087	0.007	0.321	0.120	0.179	0.097	0.182	0.101
M803Q01T	Labels	0.066	0.545	0.075	0.072	0.076	0.086	0.076	0.079	0.603	0.077	0.143	0.090	0.141	0.088
M806Q01T	Step Pattern	0.026	0.386	0.096	-0.085	0.091	-0.056	0.088	0.033	0.439	0.103	0.000	0.089	0.050	0.083
M810Q01T	Bicycles	0.017	0.393	0.100	-0.089	0.096	-0.074	0.088	0.032	0.511	0.135	-0.029	0.119	-0.080	0.119
M810Q02T	Bicycles	0.029	0.457	0.085	-0.020	0.084	0.007	0.077	0.040	0.512	0.113	0.040	0.099	0.001	0.096
M810Q03T	Bicycles	0.071	0.375	0.052	0.015	0.055	0.031	0.045	0.061	0.349	0.062	0.006	0.058	0.011	0.052
M828Q01	Carbon Dioxide	0.023	0.323	0.079	-0.054	0.070	-0.034	0.069	0.035	0.399	0.089	0.018	0.082	0.032	0.078
M828Q02	Carbon Dioxide	0.013	0.258	0.092	-0.065	0.088	-0.039	0.083	0.029	0.392	0.101	0.080	0.090	0.099	0.089
M828Q03	Carbon Dioxide	0.021	0.308	0.072	0.044	0.104	0.084	0.088	0.018	0.291	0.081	-0.031	0.071	0.003	0.070
M833Q01T	Seeing the tower	0.008	0.202	0.079	-0.117	0.070	-0.117	0.075	0.008	0.192	0.087	-0.054	0.078	-0.090	0.080

Table A6. Relationships between student responses to problem-solving items in PISA 2003 and student qualifications measured in 2007 and 2010

Unit item code	Item Name	2007						2010					
		Simple regression			After accounting for overall performance		After accounting for overall performance, gender, SES and immigrant background	Simple regression			After accounting for overall performance		After accounting for overall performance, gender, SES and immigrant background
		Expl var	Coef.	SE	Coef.	SE	Coef.	SE	Expl var	Coef.	SE	Coef.	SE	Coef.	SE
X402Q01T	Library System	0.027	0.559	0.111	0.160	0.123	0.099	0.109	0.029	0.559	0.142	0.177	0.168	0.137	0.144
X402Q02T	Library System	0.072	0.218	0.024	0.064	0.028	0.039	0.028	0.051	0.182	0.027	0.056	0.030	-0.007	0.034
X412Q01	Design by Numbers	0.034	0.392	0.066	0.016	0.060	-0.005	0.058	0.023	0.314	0.073	0.002	0.068	-0.036	0.067
X412Q02	Design by Numbers	0.042	0.428	0.063	-0.021	0.066	-0.046	0.060	0.036	0.391	0.069	0.012	0.074	-0.011	0.075
X412Q03	Design by Numbers	0.041	0.233	0.041	-0.003	0.045	0.013	0.043	0.048	0.250	0.047	0.038	0.052	0.070	0.052
X414Q01	Course Design	0.027	0.184	0.032	-0.058	0.035	-0.021	0.033	0.016	0.139	0.043	-0.072	0.042	-0.031	0.037
X415Q01T	Transit System	0.043	0.305	0.042	0.054	0.046	0.031	0.045	0.040	0.291	0.050	0.076	0.054	0.039	0.053
X417Q01	Children's Camp	0.073	0.375	0.042	0.054	0.044	0.060	0.041	0.057	0.332	0.052	0.048	0.049	0.043	0.044
X423Q01T	Freezer	0.014	0.252	0.072	-0.095	0.068	-0.093	0.063	0.017	0.272	0.081	-0.031	0.077	-0.026	0.075
X423Q02T	Freezer	0.029	0.355	0.065	-0.018	0.063	0.013	0.061	0.022	0.310	0.072	-0.017	0.072	-0.015	0.067
X430Q01	Energy Needs	0.004	0.232	0.115	-0.097	0.097	-0.104	0.088	0.010	0.338	0.142	-0.091	0.126	-0.082	0.114
X430Q02	Energy Needs	0.084	0.325	0.039	0.054	0.043	0.031	0.042	0.082	0.320	0.041	0.092	0.048	0.062	0.047
X601Q01T	Cinema Outing	0.060	0.374	0.059	0.050	0.056	0.050	0.049	0.033	0.267	0.063	-0.027	0.061	-0.042	0.055
X601Q02	Cinema Outing	0.026	0.489	0.118	0.001	0.110	0.005	0.096	0.039	0.551	0.128	0.063	0.129	0.037	0.110
X602Q01	Holiday	0.020	0.293	0.066	-0.010	0.064	0.027	0.063	0.026	0.334	0.074	0.065	0.080	0.107	0.077
X602Q02	Holiday	0.040	0.217	0.036	-0.044	0.043	-0.033	0.042	0.038	0.209	0.042	0.002	0.053	-0.045	0.047
X603Q01	Irrigation	0.017	0.365	0.087	-0.007	0.090	0.041	0.093	0.006	0.218	0.117	-0.090	0.113	-0.034	0.111
X603Q02T	Irrigation	0.038	0.432	0.075	0.048	0.081	0.077	0.078	0.012	0.246	0.087	-0.121	0.086	-0.040	0.083
X603Q03	Irrigation	0.028	0.399	0.082	-0.058	0.094	-0.020	0.091	0.020	0.330	0.095	-0.089	0.111	-0.025	0.101

Table A7. Relationship between student responses to items of different formats in PISA 2003 and student qualifications measured in 2007 and 2010: three regression models

Item format	2007				2010			
	Simple regression	After accounting for overall performance (1st plausible value)		After accounting for overall performance, gender, socio-economic and immigrant background	Simple regression	After accounting for overall performance (1st plausible value)		After accounting for overall performance, gender, socio-economic and immigrant background
	Coef.	Coef.		Coef.	Coef.	Coef.		Coef.
		After accounting for mathematics				After accounting for mathematics		
Closed constructed response	0.216***	-0.006		0.004	0.192***	-0.011		-0.011
Complex multiple choice	0.228***	-0.004		0.013	0.207***	0.004		0.015
Multiple choice	0.349***	0.070**		0.085***	0.295***	0.045		0.056*
Open constructed response	0.228***	-0.009		0.011	0.199***	-0.006		-0.013
Short response	0.306***	0.045*		0.059**	0.289***	0.057*		0.059*
		Accounting for problem-solving performance				Accounting for problem-solving performance		
Closed constructed response	0.216***	0.025		0.035	0.192***	0.013		0.011
Complex multiple choice	0.228***	0.033		0.046*	0.207***	0.024		0.033
Multiple choice	0.349***	0.095***		0.110***	0.295***	0.060*		0.069**
Open constructed response	0.228***	0.02		0.036	0.199***	0.015		0.005
Short response	0.306***	0.081***		0.094***	0.289***	0.082**		0.081**

Notes: ***, **, * denotes statistical significance at the 1%, 5% and 10% level, respectively.

Figure A1. Relationship between student responses to problem-solving items in PISA 2003 and student qualifications measured in 2007 and 2010

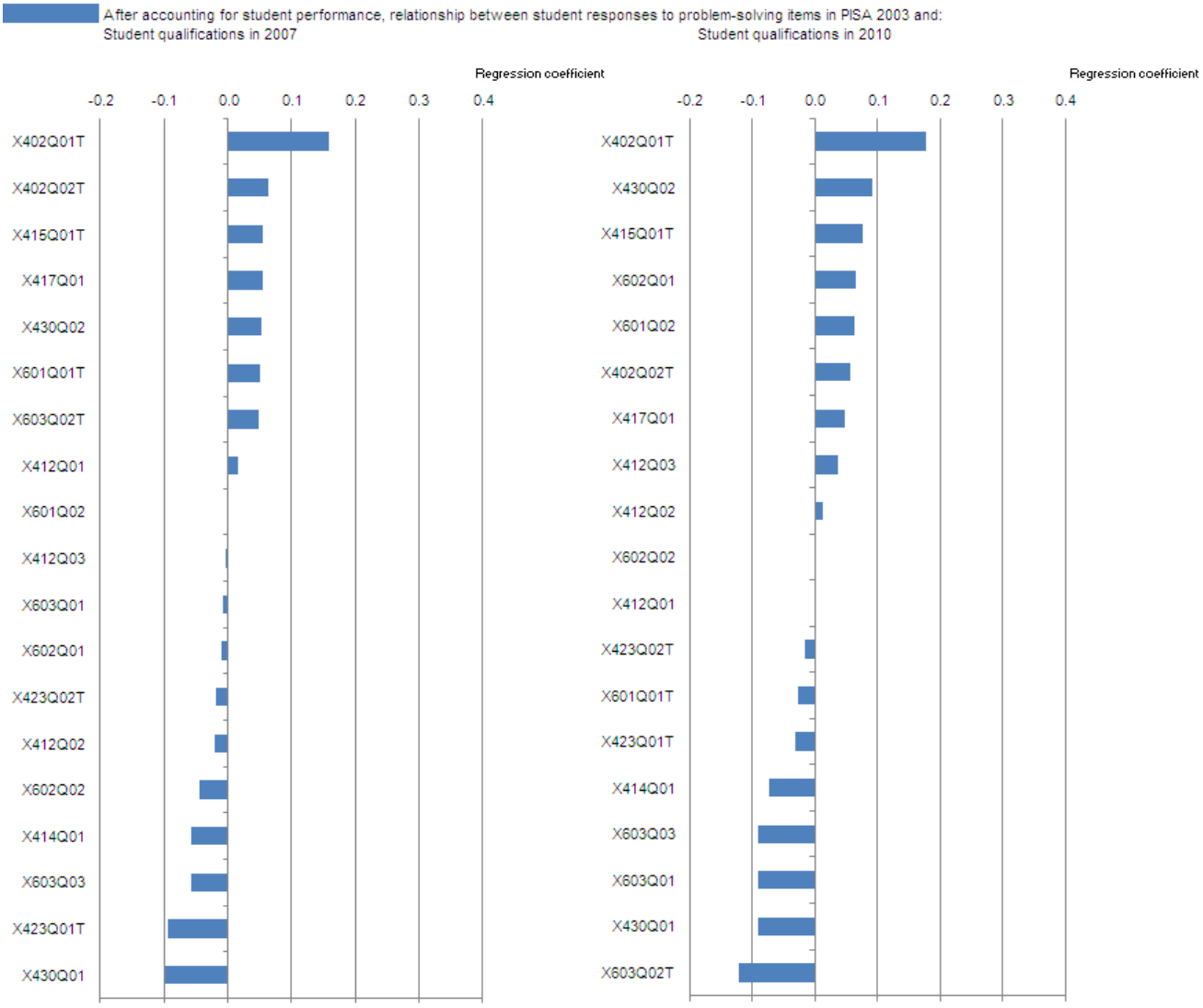


Figure A2. Correlation between the level of difficulty of a problem-solving item and the item's predictive power

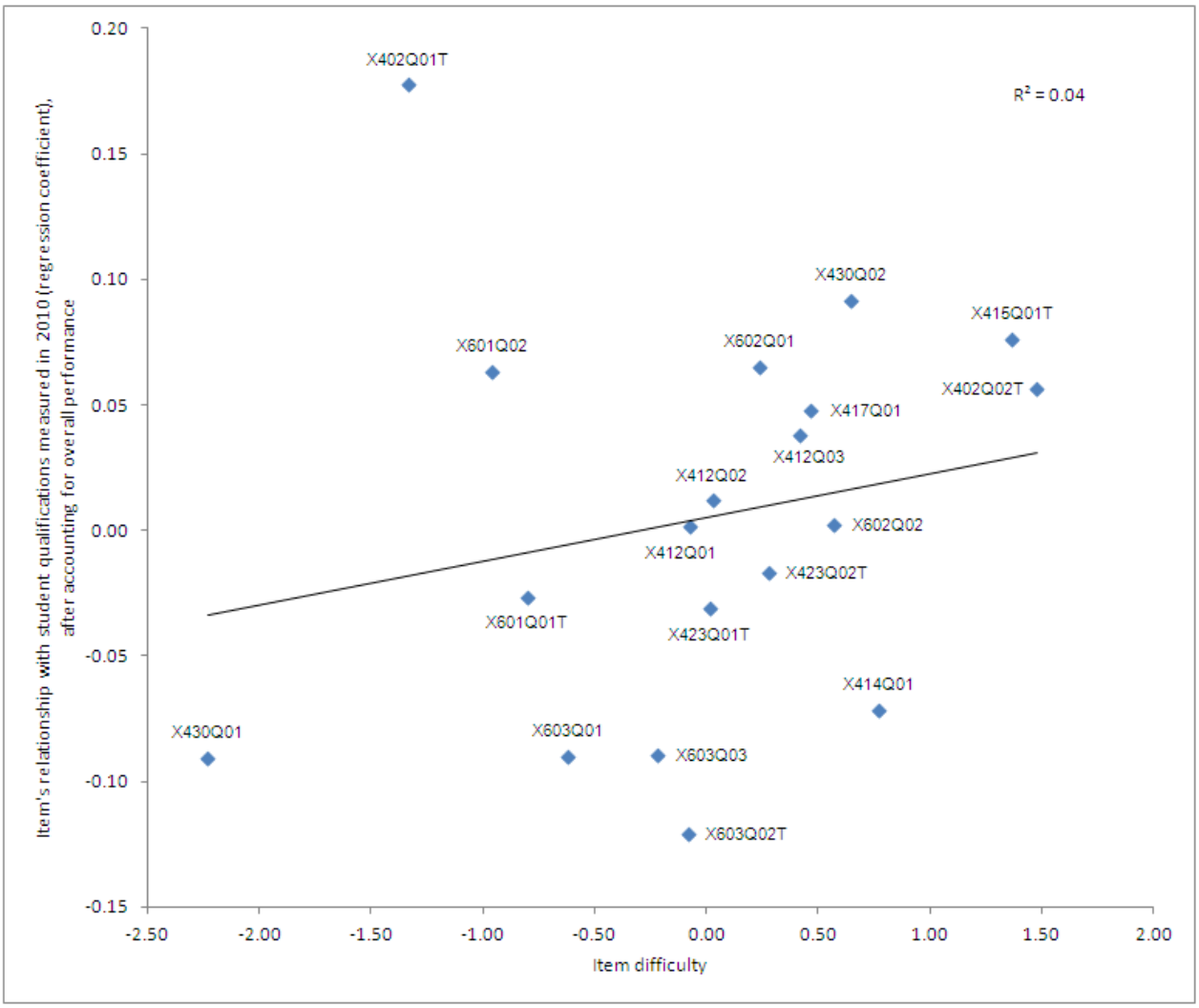
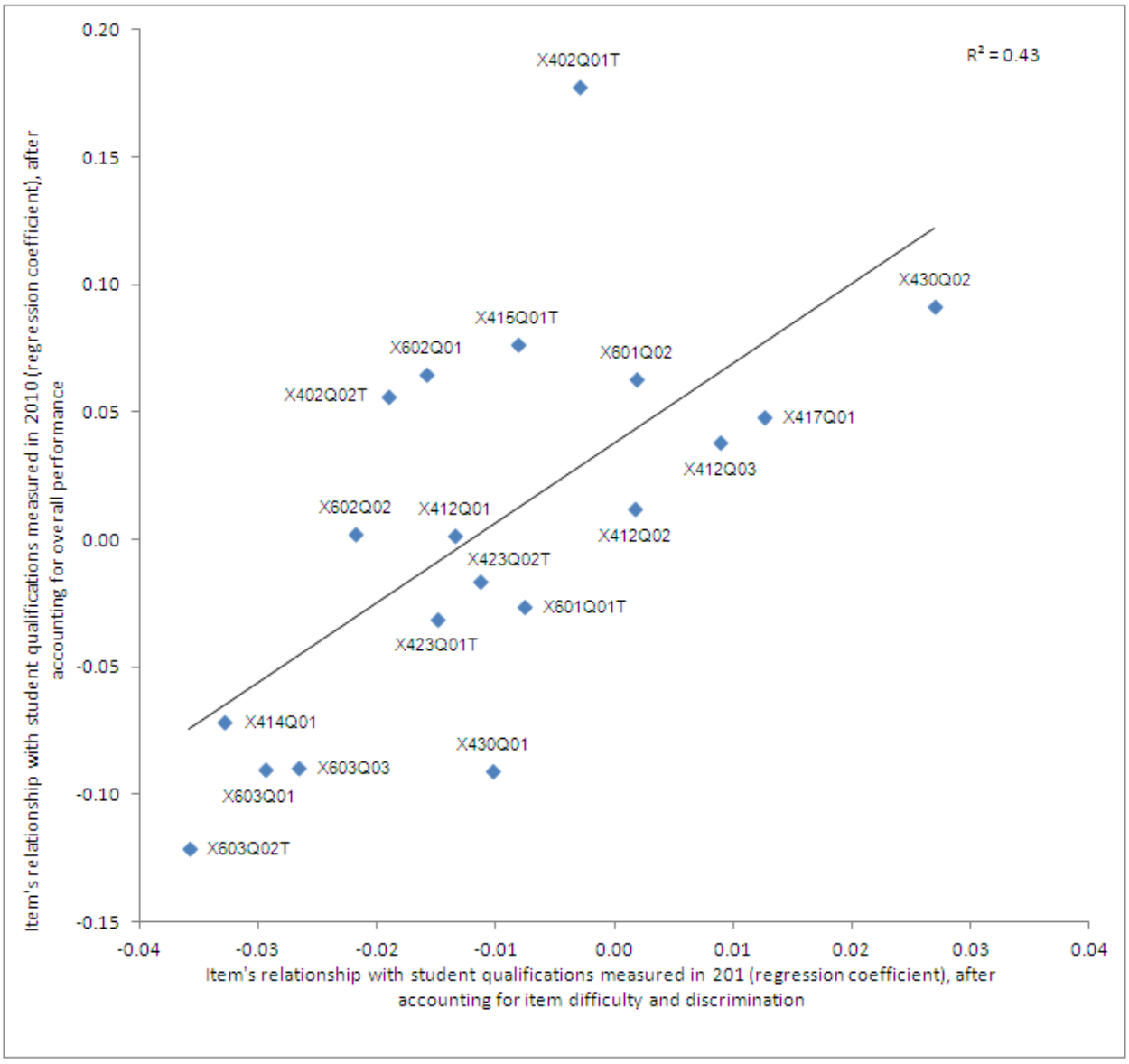


Figure A3. Correlation between two different measures of the predictive power of a problem-solving item



ANNEX B

49. This annex presents the following mathematics items:

- M555Q02T
- M402Q01 and M402Q02
- M438Q01 and M438Q02
- M467Q01

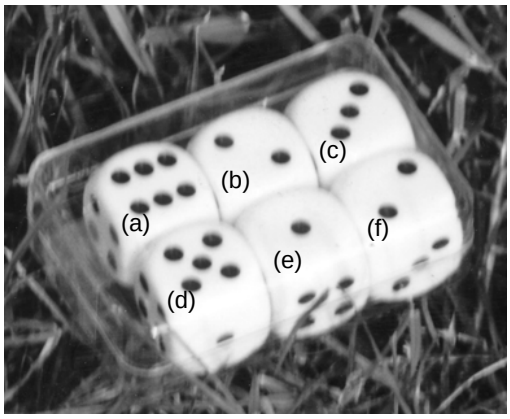
50. As shown in the columns "New classification" in Table A1 in Annex A, the item M555Q02T, which is in the unit of CUBES, is classified as *spatial reasoning* and *procedural knowledge.* The item M402Q01, which is in the unit of INTERNET RELAY CHAT, is classified as *measurement* and *procedural knowledge*. The item M402Q02, which is in the unit of INTERNET RELAY CHAT, is classified as *measurement; procedural knowledge; systematic, sequential, or strategic reasoning; thinking through a multi-step solution process* (contributes to the cognitive load/demand for an item)*; using reverse thinking/flexible thinking* (present in an item but does not determine its cognitive load/demand)*;* and *making sense of non-routine situations* (contributes to the cognitive load/demand for an item). The item M438Q01, which is in the unit of EXPORTS, is classified as *using a graph* (interpreting a display)*; using multiple graphs* (multiple-graph organisers)*;* and *procedural knowledge*. The item M438Q02, which is in the unit of EXPORTS, is classified as *thinking across multiple representations; using a graph* (reading a display)*; using multiple graphs* (multiple-graph organisers)*; procedural knowledge;* and *thinking through a multi-step solution process* (contributes to the cognitive load/demand for an item). The item M467Q01, which is in the unit of COLOURED CANDIES, is classified as *probabilistic reasoning; rates, ratios, proportions, and percentages; using a graph* (reading a display)*; using multiple graphs* (one-graph organiser)*; procedural knowledge;* and *thinking through a multi-step solution process* (present in an item but does not determine its cognitive load/demand).

M555Q02T: CUBES

Question 1

In this photograph you see six dice, labelled (a) to (f). For all dice there is a rule:

The total number of dots on two opposite faces of each die is always seven.



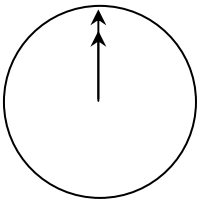
Write in each box the number of dots on the **bottom** face of the dice corresponding to the photograph.

(a)	(b)	(c)
(d)	(e)	(f)

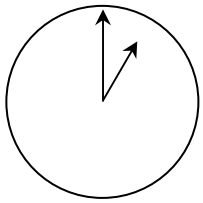
M402Q01 AND M402Q02: INTERNET RELAY CHAT

Mark (from Sydney, Australia) and Hans (from Berlin, Germany) often communicate with each other using “chat” on the Internet. They have to log on to the Internet at the same time to be able to chat.

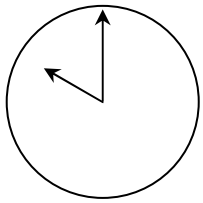
To find a suitable time to chat, Mark looked up a chart of world times and found the following:



Greenwich 12 Midnight



Berlin 1:00 AM



Sydney 10:00 AM

Question 1

At 7:00 PM in Sydney, what time is it in Berlin?

Answer: .....

Question 2

Mark and Hans are not able to chat between 9:00 AM and 4:30 PM their local time, as they have to go to school. Also, from 11:00 PM till 7:00 AM their local time they won’t be able to chat because they will be sleeping.

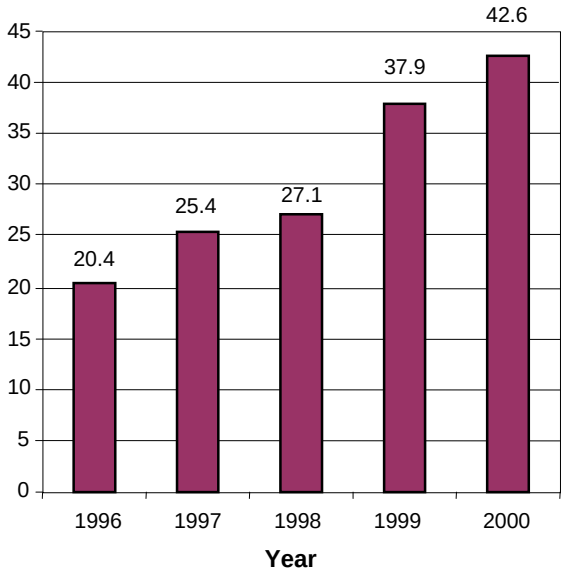
When would be a good time for Mark and Hans to chat? Write the local times in the table.

Place	Time
Sydney	
Berlin	

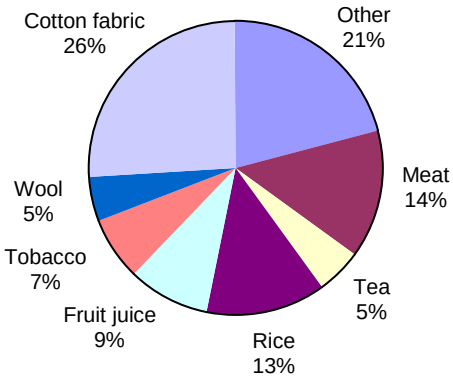
M438Q1 AND M438Q2: EXPORTS

The graphics below show information about exports from Zedland, a country that uses zeds as its currency.

Total annual exports from Zedland in millions of zeds, 1996-2000



Distribution of exports from Zedland in 2000



Question 1

What was the total value (in millions of zeds) of exports from Zedland in 1998?

Answer: .....

Question 2

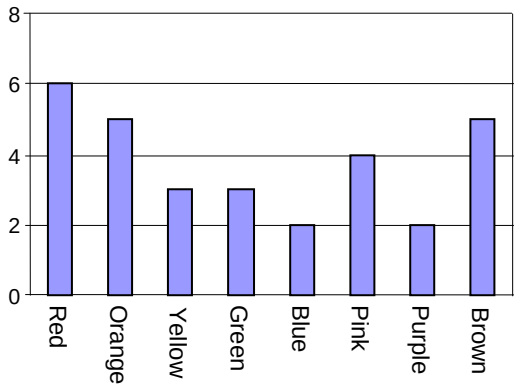
What was the value of fruit juice exported from Zedland in 2000?

A 1.8 million zeds.
B 2.3 million zeds.
C 2.4 million zeds.
D 3.4 million zeds.
E 3.8 million zeds.

M467Q01: COLOURED CANDIES

Question 1

Robert's mother lets him pick one candy from a bag. He can't see the candies. The number of candies of each colour in the bag is shown in the following graph.



What is the probability that Robert will pick a red candy?

A 10%
B 20%
C 25%
D 50%

THE OECD EDUCATION WORKING PAPERS SERIES ON LINE

The OECD Education Working Papers Series may be found at:

- The OECD Directorate for Education website: www.oecd.org/edu/workingpapers
- Online OECD-ilibrary: http://www.oecd-ilibrary.org/education/oecd-education-working-papers_19939019
- The Research Papers in Economics (RePEc) website: www.repec.org

If you wish to be informed about the release of new OECD Education working papers, please:

- Go to www.oecd.org
- Click on “My OECD”
- Sign up and create an account with “My OECD”
- Select “Education” as one of your favourite themes
- Choose “OECD Education Working Papers” as one of the newsletters you would like to receive

For further information on the OECD Education Working Papers Series, please write to: edu.contact@oecd.org.