



OECD Education Working Papers No. 151

Report on the relative
strengths and weaknesses
of the United States in PISA
2012 mathematics

**Janina Krawitz,
Kay Achmetli,
Werner Blum,
Sebastian Vogel,
Michael Besser**

<https://dx.doi.org/10.1787/bc544d1e-en>

DIRECTORATE FOR EDUCATION AND SKILLS

REPORT ON THE RELATIVE STRENGTHS AND WEAKNESSES OF THE UNITED STATES IN
PISA 2012 MATHEMATICS

OECD Education Working Paper No. 151

Janina Krawitz, Kay Achmetli, Werner Blum, Sebastian Vogel, Michael Besser - University of Kassel

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

Contact: J. Krawitz, krawitz@uni-muenster.de; K. Achmetli, kay.achmetli@uni-muenster.de; Dr. W. Blum, blum@mathematik.uni-kassel.de; S. Vogel, svogel@uni-kassel.de; Dr. M. Besser, besser@leuphana.de

JT03407097

Complete document available on OLIS in its original format

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



OECD EDUCATION WORKING PAPERS SERIES

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed herein are those of the author(s).

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcome, and may be sent to the Directorate for Education and Skills, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

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

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgement of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org.

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

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

www.oecd.org/edu/workingpapers

Copyright © OECD 2016

ACKNOWLEDGEMENTS

The authors would like to thank the OECD for providing the data on which all analyses were based as well as Michael Davidson and Francesco Avvisati for their helpful feedback and comments, and Emily Groves and Camilla Lorentzen for the editorial and administrative support provided.

ABSTRACT

This paper aims to investigate the performance of the students in the United States in all 84 mathematics items that were administered in the United States as part of the PISA 2012 assessment. It compares the performance of the United States with the OECD average and with the performance of five reference countries/economies that were ranked higher on the PISA scale. The analysis reveals specific relative strengths and weaknesses of the 15-year-olds in the United States, referring to items in which they performed unexpectedly well or unexpectedly badly compared to their overall distance from the OECD average or from the reference countries/economies. On that basis, certain patterns – that means certain clusters – of items with similar cognitive requirements, are identified. There are seven altogether, three for strengths and four for weaknesses of the US students. An analysis of student solutions illustrates and further clarifies these strengths and weaknesses. The results show that the relative strengths are mostly revealed in easy items, whereas the relative weaknesses are mostly reflected in particularly demanding items.

RÉSUMÉ

Cette étude s'intéresse aux performances des élèves aux États-Unis dans l'ensemble des 84 items de mathématiques qui ont été administrés aux États-Unis dans le cadre de l'évaluation PISA 2012. Elle compare les performances des États-Unis à la moyenne de l'OCDE et aux performances de cinq pays/économies de référence qui ont obtenu un meilleur classement à l'échelle du PISA. L'analyse révèle les forces et les faiblesses relatives spécifiques des élèves de 15 ans aux États-Unis, en se reportant aux items où ils ont obtenu des résultats les uns meilleurs et les autres pires que ceux auxquels on s'attendait, par rapport à leur classement global d'après la moyenne de l'OCDE ou les pays ou économies de référence. C'est sur cette base que sont relevés certains schémas, à savoir certains clusters, d'items présentant des exigences cognitives similaires. Il en existe sept au total, trois pour les forces et quatre pour les faiblesses des élèves américains. Une analyse des solutions trouvées par les élèves illustre et précise davantage ces forces et ces faiblesses. Les résultats montrent que les forces relatives se manifestent surtout au niveau des items faciles, tandis que les faiblesses relatives se reflètent surtout au niveau des items particulièrement exigeants.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	3
ABSTRACT	4
RÉSUMÉ.....	4
1. Introduction.....	7
2. Selection of conspicuous items.....	9
3. Patterns.....	28
4. Analyses of illustrating items.....	30
5. Analysis of student solutions	37
6. Summary	43
REFERENCES	45
ANNEX: DIF STATISTICS	46

Tables

Table 1.	Average solution rates in PISA 2012 mathematics (%)	7
Table 2.	Relative strengths of the United States compared to the average of all OECD countries (%).....	10
Table 3.	Relative weaknesses of the United States compared to the average of all OECD countries (%).....	11
Table 4.	Conspicuous items out of Table 2 compared to the reference countries/economies (%)	12
Table 5.	Conspicuous items out of Table 3 compared to the reference countries/economies (%)	13
Table 6.	Relative strengths of the United States compared to the reference countries/economies (%)...	14
Table 7.	Relative weaknesses of the United States compared to the reference countries/economies (%)	15
Table 8.	Relative strengths of the United States compared to the average of all OECD countries (in logits)	16
Table 9.	Relative weaknesses of the United States compared to the average of all OECD countries (in logits)	17
Table 10.	Conspicuous items out of Table 8 compared to the reference countries/economies (in logits).	19
Table 11.	Conspicuous items out of Table 9 compared to the reference countries/economies (in logits).	20
Table 12.	Relative strengths of the United States compared to the reference countries/economies (in logits)	21
Table 13.	Relative weaknesses of the United States compared to the reference countries/economies (in logits)	22
Table 14.	DIF-statistics for the items that are conspicuous in the logit approach	23
Table 15.	Items showing a large DIF in comparison to Asian countries/economies (negatively)	24

Table 16.	Items showing a large DIF in comparison to Asian countries/economies (positively)	25
Table 17.	Relative strengths appearing for the first time in DIF-analysis	25
Table 18.	Relative weaknesses appearing for the first time in DIF-analyses	25
Table 19.	List of conspicuous items - relative strengths.....	26
Table 20.	List of conspicuous items - relative weaknesses	27
Table 21.	Overview of the Waldtest statistics for DIF-detection sorted by size	46

Figures

Figure 1.	A model of mathematical literacy in practice	8
Figure 2.	Bubble chart of	18
Figure 3.	Bubble chart of	18
Figure 4.	Charts (PM918Q01)	31
Figure 5.	Drip rate (PM903Q03).....	33
Figure 6.	Revolving door 01 (PM995Q01).....	34
Figure 7.	Revolving door 02 (PM995Q02).....	36
Figure 8.	Revolving door 01 solution	38
Figure 9.	Revolving door 01 solution	39
Figure 10.	Revolving door 01 solution	40
Figure 11.	Revolving door 02 solution	41
Figure 12.	Revolving door 02 solution	41
Figure 13.	Revolving door 02 solution	42
Figure 14.	Revolving door 02 solution	42
Figure 15.	Revolving door 02 solution	43

1. Introduction

Students from the United States performed below average in mathematics in all cycles of the Programme for International Student Assessment (PISA). With publication of the PISA 2012 results, with mathematics as the major discipline, it seems relevant to analyse the results for the United States in more detail: are 15-year-olds in the United States below average uniformly across all topics and all competencies, or do US students have special strengths and weaknesses compared to the OECD average and, in particular, compared to the countries/economies that perform substantially higher in PISA than the United States?

Let us look at the US performance in PISA 2012 mathematics. The United States has, on average, across all 84 items administered in the United States, a significantly lower solution rate than the OECD average. The US solution rate is 43.78%, the average of all OECD countries is 47.47%. In only 18 items out of 84 did US students reach a higher solution rate than the OECD average. Other countries/economies with which the United States has to compete economically reached much higher averages in PISA mathematics. Five countries/economies have been chosen for comparisons with the United States: two top performing Asian systems (Korea and Shanghai, China), two European countries significantly above the OECD average (Germany and the Netherlands), and a neighbouring American country, Canada. These five countries/economies will be called the “reference countries/economies”. Table 1 shows the averages of these reference countries/economies in PISA 2012 mathematics (percentage).

Table 1. Average solution rates in PISA 2012 mathematics (%)

United States	OECD average	Canada	Germany	Netherlands	Korea	Shanghai, China
43.78	47.47	52.16	51.34	53.50	58.76	69.44

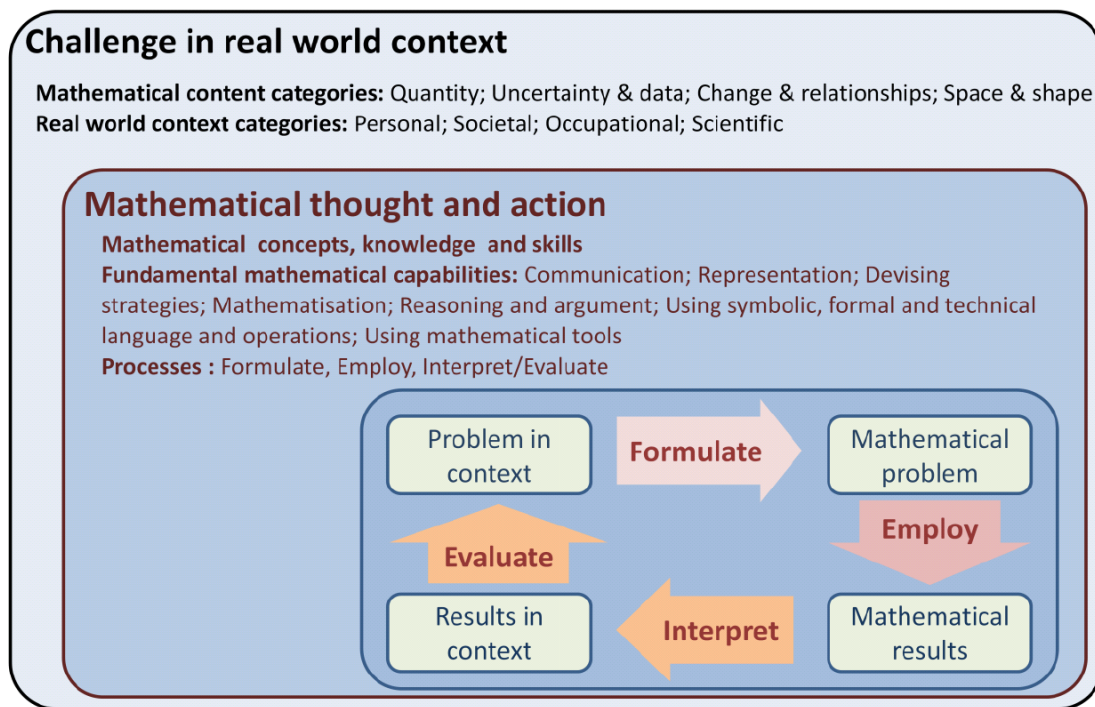
Source: Computed using the open source data from OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.

What is PISA mathematics about? The conceptual core of PISA mathematics is *mathematical literacy*. According to the PISA 2012 mathematics framework (see OECD, 2013: 25) this concept means:

...an individual’s capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognize the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens.

In short, mathematical literacy describes one’s capacity to use mathematics in a well-founded manner in order to solve real world problems, where “real world” means “...the physical, social and mental world...” (Freudenthal, 1983), including the mathematical world itself. Figure 1 (OECD, 2013a: 26) shows the processes involved when solving real world problems by means of mathematics.

Figure 1. A model of mathematical literacy in practice



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

The processes “Formulate → Employ → Interpret → Evaluate” form the so-called *modelling cycle*. Underlying these processes are certain competencies (Niss and Højgaard, 2011) called “fundamental mathematical capabilities” in the PISA framework. Seven such competencies are distinguished:

- communicating
- representing
- devising strategies
- mathematising
- reasoning and arguing
- using symbolic and formal language
- using tools.

The mathematical contents are organised into four categories: 1) quantity; 2) change and relationships; 3) space and shape; and 4) uncertainty and data. These categories are not to be confused with the traditional mathematical content areas for 15-year-olds. However, there are certain connections that are stronger than others: quantity with arithmetics; change and relationships with algebra; space and shape with geometry; and uncertainty and data with probability and statistics.

This report will look at the performance of the US students in all 84 mathematics items that were administered in the United States in PISA 2012. The US performance will be compared both with the OECD average and with the performance of the five reference countries/economies, and the specific relative strengths and relative weaknesses of the US 15-year-olds will be identified; that is, items where the US students performed unexpectedly well or unexpectedly badly compared to their overall distance from the OECD average or from the reference countries/economies. This will be done in Section 2. Altogether, there are 33 such so-called “conspicuous items” – 16 for strengths and 17 for weaknesses. It will turn out that the relative strength items are mostly easy items, whereas the relative weakness items are much more demanding. On that basis, in Section 3, certain patterns will be identified – that means certain clusters of items with similar cognitive requirements, seven altogether, three for strengths and four for weaknesses of the US students. Section 4 will analyse some of the items from these patterns more deeply in order to make the clusters better understandable. An analysis of student solutions in Section 5 will round off the analyses of strengths and weaknesses. A short summary in Section 6 will complete the report.

2. Selection of conspicuous items

Different approaches will be followed in order to identify relative strengths and relative weaknesses of the United States compared to the average of all OECD countries and compared to the five reference countries/economies Canada; Germany; Korea; the Netherlands; and Shanghai, China. To figure out the *conspicuous items* that display strengths and weaknesses, solution rates shown in percentages (Section 2.1) will be examined first and then logits (Section 2.2), and these analyses will be supported with a DIF-analysis (Section 2.3). The first approach with differences in solution rates will be rather “naïve” and exploratory, whereas for the main approach differences in logits will be compared.

2.1 First exploratory approach – solution rates

2.1.1 Comparison with the OECD average

First the absolute differences between the United States solution rates and the average solution rates of all OECD countries (rates in percentages) will be compared. The significant items will be identified, meaning those with a difference of at least one standard deviation (both sides).¹ The advantage of this approach is the easily understandable and transparent procedure using “percent correct”.

1. This standard deviation is 4.96, and the average of all differences is -3.69, both in percentages. The resulting threshold values are, therefore, 1.27 and -8.65, respectively.

Relative strengths:

Table 2. Relative strengths of the United States compared to the average of all OECD countries (%)

Item ID	Item Name	Solution rate United States	avg. solution. rate of all OECD countries	Difference	Content	Process
PM998Q02	Bike rental	80.71	71.59	9.12	Change and relationships	Interpret
PM420Q01T	Transport	57.82	50.03	7.79	Uncertainty and data	Interpret
PM446Q01	Thermometer cricket	74.74	68.59	6.14	Change and relationships	Formulate
PM954Q01	Medicine doses	70.33	65.42	4.91	Change and relationships	Employ
PM918Q01	Charts	91.93	87.27	4.66	Uncertainty and data	Interpret
PM903Q03	Drip rate	29.92	25.72	4.21	Change and relationships	Employ
PM909Q01	Speeding fines	92.91	89.35	3.56	Quantity	Interpret
PM915Q01	Carbon tax	43.56	40.20	3.37	Uncertainty and data	Employ
PM411Q02	Diving	48.94	45.73	3.21	Uncertainty and data	Interpret
PM828Q02	Carbon dioxide	58.68	55.97	2.71	Uncertainty and data	Employ
PM982Q02	Employment data	33.19	30.74	2.45	Uncertainty and data	Employ
PM155Q01	Population pyramids	69.39	67.68	1.71	Change and relationships	Interpret
PM955Q01	Migration	73.54	72.13	1.40	Uncertainty and data	Interpret
PM828Q01	Carbon dioxide	29.87	28.48	1.39	Change and relationships	Employ
PM943Q01	Arches	51.17	50.04	1.12	Change and relationships	Formulate
PM906Q02	Crazy ants	42.93	42.15	0.78	Quantity	Employ
PM905Q01T	Tennis balls	78.25	77.71	0.54	Quantity	Interpret
PM155Q02D	Population pyramids	62.09	61.59	0.51	Change and relationships	Employ

List of items in order of differences from high to low in which the United States performed significantly better (first fourteen items) or at least better (the last four items) than the OECD average (comparison of percentages).

Source: Computed using the open source data from OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.

This method reveals 14 items for which the differences between the solution rates of the United States and the OECD average are significantly higher than the average difference. Four more items are listed here where the US solution rates are still bigger than the OECD average rates.

It is remarkable that there is no item among those fourteen items from the content area “Space and shape” and only one item from “Quantity”, thus 13 items are either assigned to “Change and relationships” or to “Uncertainty and data”. Equally remarkable is that only one item is assigned to the process category “Formulate”, whereas seven items stem from the category “Interpret” and six items from “Employ”. The range between the US solution rates and the OECD average is not that big in comparison to those items in which the United States performed worse (see Table 3, below). With differences of 9.12 percentage points for the Item “Bike rental 02”, 7.79 percentage points for “Transport 01” and 6.14 percentage points for “Thermometer cricket 01” there are only three items for which the difference of the solution rates is higher than 5 percentage points. Most of the items here are easy items with relatively high average solution rates.

Relative weaknesses:

Table 3. Relative weaknesses of the United States compared to the average of all OECD countries (%)

Item ID	Item Name	solution rate United States	avg. solution rate of all OECD countries	Difference	Content	Process
PM915Q02	Carbon tax	52.94	68.26	-15.32	Change and relationships	Employ
PM955Q02	Migration	19.61	34.23	-14.62	Uncertainty and data	Interpret
PM305Q01	Map	46.45	60.37	-13.91	Space and shape	Employ
PM924Q02	Sauce	51.32	63.48	-12.16	Quantity	Formulate
PM905Q02	Tennis balls	38.70	50.07	-11.36	Quantity	Employ
PM995Q01	Revolving door	46.96	57.71	-10.75	Space and shape	Employ
PM906Q01	Crazy ants	50.70	60.67	-9.96	Quantity	Employ
PM949Q01T	Roof truss design	57.68	67.54	-9.86	Space and shape	Formulate
PM034Q01T	Bricks	32.69	42.40	-9.71	Space and shape	Formulate
PM464Q01T	The Fence	14.08	23.67	-9.59	Space and shape	Formulate
PM923Q01	Sailing ships	50.09	59.50	-9.41	Quantity	Employ
PM406Q01	Running tracks	16.69	25.63	-8.94	Space and shape	Employ
PM800Q01	Computer game	79.49	88.41	-8.92	Quantity	Employ

List of items in order of differences from high to low (absolute values) in which the United States performed significantly worse than the OECD average (comparison of percentages).

Source: Computed using the open source data from OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.

This method reveals thirteen items for which the differences between the solution rates of the United States and the OECD average are significantly lower than the average difference. It is noticeable that nearly half of the listed items belong to the content area “Space and shape”; there is only one item from the content area “Change and relationships” and also only one from “Uncertainty and data”. Furthermore, there is only one item labelled in the process category as “Interpret”, whereas eight items are assigned to “Employ” and four items to “Formulate”. The differences range from 8.92 up to 15.32 percentage points. In addition most of the items here are relatively easy. However, this is also due to the method of comparing solution rates since for difficult items with low solution rates the absolute differences naturally tend to be smaller.

It is interesting that the two items with the biggest difference of solution rates, “Carbon tax 02” and “Migration 02”, are both from tasks in which the United States performed better than the OECD average in the first items of these tasks (“Carbon tax 01” and “Migration 01”). So the context itself seems not to be responsible for these differences. A closer look is required to reveal possible reasons for the big differences between the first and second part of these tasks. In Section 3, these items were assigned to certain patterns in order to understand this behaviour.

2.1.2 Comparison with the reference countries/economies

To classify the performance of the United States in more detail, we will take the items identified so far (Table 2 and Table 3) and compare the United States with the five *reference countries/economies* for those items (solution rates as a percentage).

Relative strengths:

Table 4. Conspicuous items out of Table 2 compared to the reference countries/economies (%)

Item ID	Item Name	solution rate United States	avg. solution rate of all OECD countries	solution rate Canada	solution rate Germany	solution rate Netherlands	solution rate Korea	solution rate Shanghai (China)
PM998Q02	Bike rental	80.71	71.59	84.93	76.06	89.07	80.06	87.63
PM420Q01T	Transport	57.82	50.03	64.12	50.70	57.67	28.19	60.67
PM446Q01	Thermometer cricket	74.74	68.59	77.93	65.00	73.19	79.33	88.07
PM954Q01	Medicine doses	70.33	65.42	71.93	68.42	58.88	81.18	92.45
PM918Q01	Charts	91.93	87.27	86.89	87.95	90.43	90.76	92.55
PM903Q03	Drip rate	29.92	25.72	37.09	24.79	30.62	46.27	70.77
PM909Q01	Speeding fines	92.91	89.35	94.40	92.44	95.72	93.51	94.43
PM915Q01	Carbon tax	43.56	40.20	40.72	31.91	49.44	66.66	72.63
PM411Q02	Diving	48.94	45.73	51.92	49.66	62.04	56.45	71.11
PM828Q02	Carbon dioxide	58.68	55.97	62.50	59.80	69.44	66.28	76.04
PM982Q02	Employment data	33.19	30.74	35.81	28.72	30.06	44.53	37.59
PM155Q01	Population pyramids	69.39	67.68	74.31	69.58	79.92	77.51	72.66
PM955Q01	Migration	73.54	72.13	79.06	79.03	70.39	75.15	84.35
PM828Q01	Carbon dioxide	29.87	28.48	38.32	33.09	18.20	49.06	62.09

List of items in order of differences from high to low in which the United States performed significantly better than the OECD average.

Note: Items in which the United States performed better than the reference countries are highlighted in blue (comparison of percentages).

Source: Computed using the open source data from OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.

As is known, the United States' performance in PISA 2012 was, on average, substantially worse than that of the reference countries/economies. Any item in which the United States reached a higher solution rate than one of the reference countries/economies is, therefore, interesting. Table 4 contains the same items as presented in Table 2. Each item with a lower solution rate than the US rate is highlighted in blue.

In none of these items did the United States perform better than Shanghai, China. Additionally, there are three items – “Diving 02”, “Carbon dioxide 02” and “Population pyramids 01” – where the United States did not perform better than any of the reference countries/economies. Only in two items (“Charts 01” and “Transport 01”) did the United States score considerably better than three out of the five reference countries/economies. In comparison to Germany, the United States did considerably well for these items; that is, in nine out of fourteen items they performed better. Compared to the Netherlands there are six such items, to Korea three and to Canada two items in which the United States has a higher solution rate.

Relative weaknesses:

Table 5. Conspicuous items out of Table 3 compared to the reference countries/economies (%)

Item ID	Item Name	solution rate United States	avg. solution rate of all OECD countries	resolution rate Canada	solution rate Germany	solution rate Netherlands	solution rate Korea	solution rate Shanghai (China)
PM915Q02	Carbon tax	52.94	68.26	76.10	80.46	75.40	74.02	90.10
PM955Q02	Migration	19.61	34.23	34.95	36.35	48.47	56.27	57.59
PM305Q01	Map	46.45	60.37	63.47	66.34	67.05	62.46	71.95
PM924Q02	Sauce	51.32	63.48	60.64	66.49	81.32	73.22	85.13
PM905Q02	Tennis balls	38.70	50.07	57.38	56.35	61.69	63.12	80.62
PM995Q01	Revolving door	46.96	57.71	57.44	63.99	59.09	79.08	89.70
PM906Q01	Crazy ants	50.70	60.67	62.11	67.66	75.23	61.07	79.18
PM949Q01T	Roof truss design	57.68	67.54	75.97	73.06	77.77	85.38	89.75
PM034Q01T	Bricks	32.69	42.40	48.07	48.17	46.07	55.57	54.04
PM464Q01T	The fence	14.08	23.67	30.46	32.58	23.51	41.98	59.82
PM923Q01	Sailing ships	50.09	59.50	58.03	64.10	78.08	69.10	73.91
PM406Q01	Running tracks	16.69	25.63	35.52	30.90	29.79	40.12	57.83
PM800Q01	Computer game	79.49	88.41	87.28	84.64	88.33	96.45	96.88

List of items in order of differences from high to low (absolute values) in which the United States performed significantly worse than the OECD average.

Note: The largest and the smallest difference between solution rates for the United States and reference countries are highlighted in blue (comparison of percentages).

Source: Computed using the open source data from OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.

The United States has a lower solution rate of minimum 5.16 percentage points for these items; this lowest difference exists between the United States and Germany for the Item “Computer game 01” (see Table 5). The differences increase up to 45.74 percentage points; this is reached for the item “The Fence 01” in comparison with Shanghai, China. There are also big differences between the reference countries/economies; e.g. for the item “Revolving door 01”, Korea and Shanghai, China achieved much higher solution rates than the other countries/economies.

2.1.3 Comparison with the average of the reference countries/economies

New conspicuous items can be found by generating an average solution rate for each item referring only to the five reference countries/economies (see Table 6). This average is compared to the solution rate of the United States (again: the difference between the US solution rate and average solution rate of these countries/economies). “Unexpectedly large” and “unexpectedly small” differences will be looked for. More precisely: all items will be considered where the difference between those two solution rates is larger or smaller than the average of all those differences by at least one standard deviation.² “Large” differences

2. This standard deviation is 6.72, and the average is -13.26, all in percentages. The resulting threshold values are, therefore, -6.54 and -19.98, respectively.

are then all those that are positive or negative, but still significantly above the average difference. “Small” differences are all those that are significantly below the average difference.

Relative strengths:

Table 6. Relative strengths of the United States compared to the reference countries/economies (%)

Item ID	Item Name	solution rate United States	avg. solution rate of the ref. countries/ economies	solution rate Canada	solution rate Germany	solution rate Netherlands	solution rate Korea	solution rate Shanghai (China)
PM420Q01T	Transport	57.82	52.27	64.12	50.70	57.67	28.19	60.67
PM918Q01	Charts	91.93	89.72	86.89	87.95	90.43	90.76	92.55
PM982Q01	Employment data	87.23	86.50	87.76	89.08	88.66	85.34	81.65
PM909Q01	Speeding fines	92.91	94.10	94.40	92.44	95.72	93.51	94.43
PM446Q01	Thermometer cricket	74.74	76.70	77.93	65.00	73.19	79.33	88.07
PM982Q02	Employment data	33.19	35.34	35.81	28.72	30.06	44.53	37.59
PM998Q02	Bike rental	80.71	83.55	84.93	76.06	89.07	80.06	87.63
PM955Q01	Migration	73.54	77.60	79.06	79.03	70.39	75.15	84.35
PM995Q02	Revolving door	2.26	6.33	4.26	3.32	4.09	6.35	13.62
PM954Q01	Medicine doses	70.33	74.57	71.93	68.42	58.88	81.18	92.45
PM905Q01T	Tennis balls	78.25	83.46	82.53	79.57	82.84	86.18	86.18
PM155Q01	Population pyramids	69.39	74.80	74.31	69.58	79.92	77.51	72.66
PM033Q01	A view room	73.04	79.33	77.84	81.60	80.25	78.22	78.73

Note: List of items in which the difference of the solution rate between the United States and the solution rate of the reference countries (average) is significantly above the average of these differences. The new items that did not appear in the comparison with the OECD average (Table 2) are bolded (comparison of percentages). The blue shading indicates the United States performs better than the comparator country.

Source: Computed using the open source data from OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.

This method reveals thirteen items with significant differences in the solution rates of the United States compared with the average of the five reference countries/economies. Four out of these thirteen items are new: they are bolded. “Employment data 01” is the only new item where the solution rate of the United States is higher than in at least one reference country. The item “Revolving door 02” displays a lower solution rate for the United States than for every reference country and, furthermore, a lower solution rate than the average of all OECD countries. The reason why this item appears in this approach is that the solution rates are very low overall, thus the differences both to the OECD average and to the average of the reference countries/economies are, necessarily, so low that this item still appears as a relative strengths for the United States. This effect of items with very low success rates suggests that our first heuristic approach does not fit our purposes sufficiently well.

Relative weaknesses:

Table 7. Relative weaknesses of the United States compared to the reference countries/economies (%)

Item ID	Item Name	solution rate United States	avg. solution rate of the ref. countries/ economies	solution rate Canada	solution rate Germany	solution rate Netherlands	solution rate Korea	solution rate Shanghai (China)
PM955Q02	Migration	19.61	46.72	34.95	36.35	48.47	56.27	57.59
PM915Q02	Carbon tax	52.94	79.22	76.10	80.46	75.40	74.02	90.10
PM905Q02	Tennis balls	38.70	63.83	57.38	56.35	61.69	63.12	80.62
PM464Q01T	The fence	14.08	37.67	30.46	32.58	23.51	41.98	59.82
PM995Q01	Revolving door	46.96	69.86	57.44	63.99	59.09	79.08	89.70
PM949Q01T	Roof truss design	57.68	80.39	75.97	73.06	77.77	85.38	89.75
PM406Q01	Running tracks	16.69	38.83	35.52	30.90	29.79	40.12	57.83
PM924Q02	Sauce	51.32	73.36	60.64	66.49	81.32	73.22	85.13
PM909Q03	Speeding fines	28.21	49.89	40.01	44.93	47.43	55.00	62.10
PM949Q02T	Roof truss design	27.81	49.18	38.49	26.59	21.14	73.62	86.04
PM192Q01T	Containers	35.86	56.71	50.22	50.77	59.48	58.18	64.88
PM953Q03	Flu test	44.29	65.04	56.27	56.60	63.86	67.13	81.34
PM949Q03	Roof truss design	31.45	51.81	35.93	32.00	39.54	68.39	83.20

Note: List of items in which the difference of the solution rate between the United States and the solution rate of the reference countries (average) is significantly beneath the average of these differences. The new items that did not appear in the comparison with the OECD average (Table 3) are bolded. Further, the item with the largest difference to one of the reference countries is also highlighted (comparison of percentages).

Source: Computed using the open source data from OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.

There are thirteen conspicuous items, five of them are new: they are bolded. Noticeable is the big difference between Shanghai, China and the United States. The differences between these two countries/economies increase up to 58.24 percentage points, namely for the item “Roof truss design 02”. There Shanghai, China has a solution rate of 86.04%, whereas the United States only has a solution rate of 27.81%. It is interesting that, for precisely this item, the United States has even a slightly higher solution rate than Germany and the Netherlands. Obviously, this item displays a particular strength of the Asian reference countries/economies rather than a relative weakness of the United States.

2.2 Main approach – logits

The disadvantage of our first exploratory approach is that it does not reveal differences in items with overall relatively low or very high solution rates. The relationship between getting items correct and their difficulty is not linear, so comparing solution rates of 95% and 85% means something different than comparing 55% and 45%, but should be roughly the same as 15% and 5%. Therefore, it is necessary to transform the percentages into a linear metric. This can be achieved with a transformation of the percentages into *logits*. This transformation has the effect that very low and very high solution rates will be “stretched out” in comparison to solution rates close to 50%. A logit value of “0” means that the item has a solution rate of 50%. Positive logits mean higher solution rates and negative logits mean lower solution rates.

2.2.1 Comparison with the OECD average

First the logit differences between the United States and the average of all OECD countries will be compared. The significant items – meaning a difference of at least one standard deviation (both sides) – will be identified (see Table 8 and Table 9).³

Relative strengths:

Table 8. Relative strengths of the United States compared to the average of all OECD countries (in logits)

Item ID	Item name	logit United States	logit avg. all OECD countries	Difference	Content	Process
PM918Q01	Charts	2.43	1.92	0.51	Uncertainty and data	Interpret
PM998Q02	Bike rental	1.43	0.92	0.51	Change and relationships	Interpret
PM909Q01	Speeding fines	2.57	2.13	0.45	Quantity	Interpret
PM420Q01T	Transport	0.32	0.00	0.31	Uncertainty and data	Interpret
PM446Q01	Thermometer cricket	1.08	0.78	0.30	Change and relationships	Formulate
PM954Q01	Medicine doses	0.86	0.64	0.23	Change and relationships	Employ
PM903Q03	Drip rate	-0.85	-1.06	0.21	Change and relationships	Employ
PM915Q01	Carbon tax	-0.26	-0.40	0.14	Uncertainty and data	Employ
PM411Q02	Diving	-0.04	-0.17	0.13	Uncertainty and data	Interpret
PM982Q02	Employment data	-0.70	-0.81	0.11	Uncertainty and data	Employ
PM828Q02	Carbon dioxide	0.35	0.24	0.11	Uncertainty and data	Employ
PM155Q01	Population pyramids	0.82	0.74	0.08	Change and relationships	Interpret

List of items in order of differences from high to low in which the United States performed significantly better than the OECD average (comparison of logits).

Source: Computed using the open source data from OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.

This method reveals twelve conspicuous items with significant differences in the logits of the United States compared to the average of all OECD countries. None of them are new compared to the previous approach with solution rates, which is not so surprising, since these are all items where the United States scored better than the OECD average, no matter how this average is measured. However, the order has changed due to the logit transformation. In contrast to the approach with solution rates, “Migration 01” and “Carbon Dioxide 01” no longer have a significant difference.

3. This standard deviation is 0.29, and the average of all differences is -0.21. The resulting threshold values are, therefore, 0.08 and -0.50, respectively.

Relative weaknesses:

Table 9. Relative weaknesses of the United States compared to the average of all OECD countries (in logits)

Item ID	Item name	logit United States	logit avg. all OECD countries	Difference	Content	Process
PM00KQ02	Wheelchair basketball	-2.59	-1.75	-0.84	Space and shape	Formulate
PM943Q02	Arches	-3.67	-2.89	-0.79	Space and shape	Formulate
PM955Q03	Migration	-2.76	-1.99	-0.76	Uncertainty and data	Employ
PM955Q02	Migration	-1.41	-0.65	-0.76	Uncertainty and data	Interpret
PM406Q02	Running tracks	-2.33	-1.59	-0.74	Space and shape	Formulate
PM800Q01	Computer game	1.35	2.03	-0.68	Quantity	Employ
PM915Q02	Carbon tax	0.12	0.77	-0.65	Change and relationships	Employ
PM464Q01T	The fence	-1.81	-1.17	-0.64	Space and shape	Formulate
PM305Q01	Map	-0.14	0.42	-0.56	Space and shape	Employ
PM406Q01	Running tracks	-1.61	-1.07	-0.54	Space and shape	Employ
PM00GQ01	An advertising column	-2.87	-2.34	-0.53	Space and shape	Formulate
PM924Q02	Sauce	0.05	0.55	-0.50	Quantity	Formulate

List of items in order of differences from high to low (absolute values) in which the US performed significantly worse than the OECD average (comparison of logits).

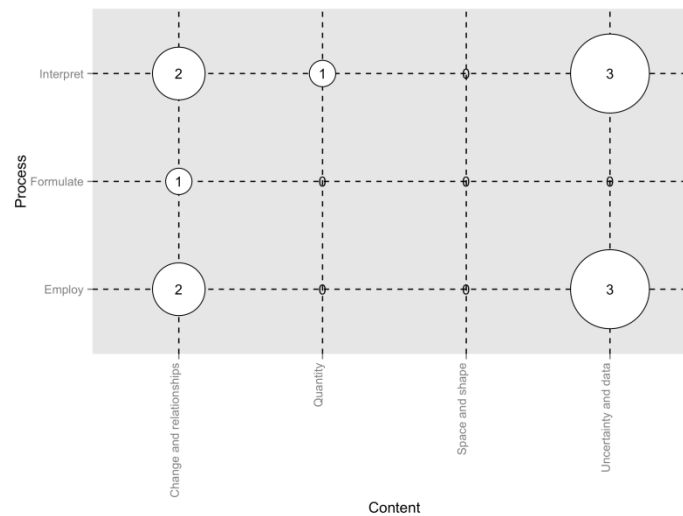
Note: The new items that did not appear in the former comparisons (Table 3 and Table 7) are in bold.

Source: Computed using the open source data from OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.

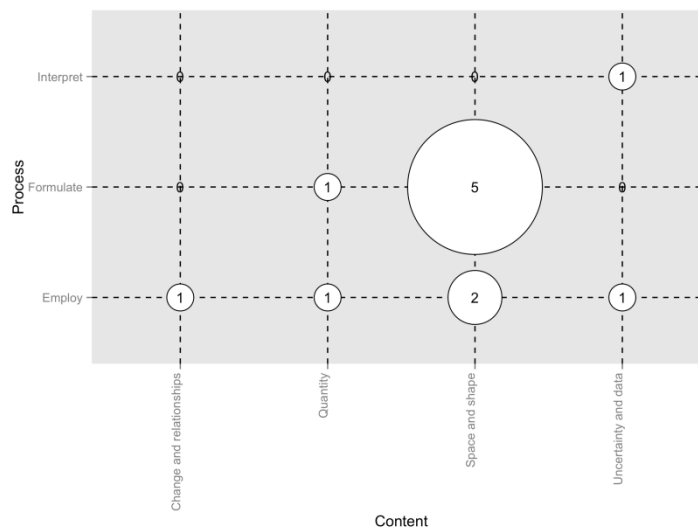
Table 9 displays twelve conspicuous items relating to the differences between the logit values of the average of all OECD countries and the United States. With this method, five new conspicuous items can be detected compared to the previous approach with solution rates, in bold. All of these new items are relatively difficult, because the logit transformation compensates the floor effect of the absolute differences in percentages. Therefore, these items could not be discovered in the solution rates approach.

2.2.2 Conspicuous items in “Content” and “Process”

In order to get a first impression of similarities and differences of the conspicuous items for the relative strengths and weaknesses of the United States, it is helpful to have a look at the item categories “Content” and “Process” (see Section 1). Therefore, the correlations between the “Contents” and the “Processes” of conspicuous items are examined in Figure 2 and Figure 3.

Figure 2. Bubble chart of “Process” and “Content” for the relative strengths

Source: Created by the authors.

Figure 3. Bubble chart of "Process" and "Content" for the relative weakness

Source: Created by the authors.

It can be seen that the overlap between these bubble charts for the strengths and weaknesses is almost negligible. Obviously, there are several different clusters. The strengths of the US students manifest themselves primarily in items that require employing or interpreting mathematical concepts or facts and are either from the mathematical content area “Change and relationships” or from “Uncertainty and data”. In contrast, the weaknesses of US students manifest themselves primarily in items that require formulating or employing mathematical concepts or facts from the mathematical content area “Space and shape”.

2.2.3 Comparison with the reference countries/economies

Again, the United States will be compared with the five reference countries/economies for all hitherto identified conspicuous items.

Relative strengths:

Table 10. Conspicuous items out of Table 8 compared to the reference countries/economies (in logits)

Item ID	Item name	logit United States	logit avg. all OECD countries	logit Canada	logit Germany	logit Netherlands	logit Korea	logit Shanghai (China)
PM918Q01	Charts	2.43	1.92	1.89	1.99	2.25	2.29	2.52
PM998Q02	Bike rental	1.43	0.92	1.73	1.16	2.10	1.39	1.96
PM909Q01	Speeding fines	2.57	2.13	2.82	2.50	3.11	2.67	2.83
PM420Q01T	Transport	0.32	0.00	0.58	0.03	0.31	-0.94	0.43
PM446Q01	Thermometer cricket	1.08	0.78	1.26	0.62	1.00	1.34	2.00
PM954Q01	Medicine doses	0.86	0.64	0.94	0.77	0.36	1.46	2.51
PM903Q03	Drip rate	-0.85	-1.06	-0.53	-1.11	-0.82	-0.15	0.88
PM915Q01	Carbon tax	-0.26	-0.40	-0.38	-0.76	-0.02	0.69	0.98
PM411Q02	Diving	-0.04	-0.17	0.08	-0.01	0.49	0.26	0.90
PM982Q02	Employment data	-0.70	-0.81	-0.58	-0.91	-0.84	-0.22	-0.51
PM828Q02	Carbon dioxide	0.35	0.24	0.51	0.40	0.82	0.68	1.16
PM155Q01	Population pyramids	0.82	0.74	1.06	0.83	1.38	1.24	0.98

List of items in order of differences from high to low in which the United States performed significantly better than the OECD average.

Note: Items in which the United States performed better than the reference countries are highlighted (comparison of logits).

Source: Computed using the open source data from OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.

Table 10 displays, of course, similar structures to those in Table 4 (shown there as “percent correct”). The order has changed due to the logit transformation.

Relative weaknesses:

Table 11. Conspicuous items out of Table 9 compared to the reference countries/economies (in logits)

Item ID	Item Name	logit United States	logit avg. all OECD countries	logit Canada	logit Germany	logit Netherlands	logit Korea	logit Shanghai (China)
PM00KQ02	Wheelchair basketball	-2.59	-1.75	-1.44	-1.31	-1.53	-1.77	-0.17
PM943Q02	Arches	-3.67	-2.89	-3.14	-2.78	-3.39	-1.42	0.04
PM955Q03	Migration	-2.76	-1.99	-1.74	-1.79	-1.25	-1.16	-0.29
PM955Q02	Migration	-1.41	-0.65	-0.62	-0.56	-0.06	0.25	0.31
PM406Q02	Running tracks	-2.33	-1.59	-1.14	-1.42	-1.21	-0.84	-0.20
PM800Q01	Computer game	1.35	2.03	1.93	1.71	2.02	3.30	3.43
PM915Q02	Carbon tax	0.12	0.77	1.16	1.42	1.12	1.05	2.21
PM464Q01T	The fence	-1.81	-1.17	-0.83	-0.73	-1.18	-0.32	0.40
PM305Q01	Map	-0.14	0.42	0.55	0.68	0.71	0.51	0.94
PM406Q01	Running tracks	-1.61	-1.07	-0.60	-0.80	-0.86	-0.40	0.32
PM00GQ01	An advertising column	-2.87	-2.34	-2.28	-2.06	-1.70	-1.72	-1.52
PM924Q02	Sauce	0.05	0.55	0.43	0.69	1.47	1.01	1.74

List of items in order of differences from high to low (absolute values) in which the United States performed significantly worse than the OECD average (comparison of logits).

Source: Computed using the open source data from OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.

2.2.4 Comparison with the average of the reference countries/economies

New conspicuous items can be found by generating an *average logit value* for each item referring only to the five reference countries/economies. This average is compared with the logits of the United States. Again, “unexpectedly large” and “unexpectedly small” differences will be looked for. More precisely: all items will be considered where the difference between those two logits (United States minus average of the reference countries/economies) is bigger or smaller than the average of all differences by at least one standard deviation.⁴

4. This standard deviation is 0.36, and the average is -0.69. The resulting threshold values are, therefore, 0.33 and -1.05, respectively.

Relative strengths:

Table 12. Relative strengths of the United States compared to the reference countries/economies (in logits)

Item ID	Item name	logit United States	logit avg. ref. countries/economies	logit Canada	logit Germany	logit Netherlands	logit Korea	logit Shanghai (China)
PM918Q01	Charts	2.43	2.19	1.89	1.99	2.25	2.29	2.52
PM420Q01T	Transport	0.32	0.08	0.58	0.03	0.31	-0.94	0.43
PM982Q01	Employment data	1.92	1.88	1.97	2.10	2.06	1.76	1.49
PM982Q02	Employment data	-0.70	-0.61	-0.58	-0.91	-0.84	-0.22	-0.51
PM446Q01	Thermometer cricket	1.08	1.25	1.26	0.62	1.00	1.34	2.00
PM909Q01	Speeding fines	2.57	2.79	2.82	2.50	3.11	2.67	2.83
PM998Q02	Bike rental	1.43	1.67	1.73	1.16	2.10	1.39	1.96
PM955Q01	Migration	1.02	1.26	1.33	1.33	0.87	1.11	1.68
PM155Q01	Population pyramids	0.82	1.10	1.06	0.83	1.38	1.24	0.98
PM155Q04T	Population pyramids	0.16	0.47	0.56	0.35	0.04	0.59	0.81
PM943Q01	Arches	0.05	0.37	0.15	0.08	-0.20	0.52	1.30

List of items in which the difference of the solution rate in logits between the United States and the average of the reference countries is significant above the average of these differences.

Note: The new items that did not appear in former comparisons (Table 2 and Table 10) are in bold. The blue shading indicates the United States performs better than the comparator country.

Source: Computed using the open source data from OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.

This method reveals four new conspicuous items, in bold (see Table 12). The item “Employment data 01” shows, remarkably enough, a positive value compared to the two Asian countries/economies, while the other three items show a positive logit value in comparison to the Netherlands. The results of “Employment data 01” might also reveal one of the rare relative weaknesses of the Asian countries/economies.

Relative weaknesses:

Table 13. Relative weaknesses of the United States compared to the reference countries/economies (in logits)

Item ID	Item name	logit United States	logit avg. ref. countries/economies	logit Canada	logit Germany	logit Netherlands	logit Korea	logit Shanghai (China)
PM943Q02	Arches	-3.67	-2.14	-3.14	-2.78	-3.67	-1.42	0.04
PM955Q03	Migration	-2.76	-1.24	-1.74	-1.79	-2.76	-1.16	-0.29
PM406Q02	Running tracks	-2.33	-0.96	-1.14	-1.42	-2.33	-0.84	-0.20
PM00KQ02	Wheelchair basketball	-2.59	-1.24	-1.44	-1.31	-2.59	-1.77	-0.17
PM464Q01T	The fence	-1.81	-0.53	-0.83	-0.73	-1.81	-0.32	0.40
PM955Q02	Migration	-1.41	-0.14	-0.62	-0.56	-1.41	0.25	0.31
PM915Q02	Carbon tax	0.12	1.39	1.16	1.42	0.12	1.05	2.21
PM949Q01T	Roof truss design	0.31	1.47	1.15	1.00	0.31	1.76	2.17
PM406Q01	Running tracks	-1.61	-0.47	-0.60	-0.80	-1.61	-0.40	0.32
PM992Q03	Spacers	-2.86	-1.73	-2.35	-2.23	-2.86	-0.98	-0.49
PM800Q01	Computer game	1.35	2.48	1.93	1.71	1.35	3.30	3.43
PM446Q02	Thermometer cricket	-2.87	-1.74	-2.33	-2.55	-2.87	-1.64	-0.36
PM995Q01	Revolving door	-0.12	0.95	0.30	0.58	-0.12	1.33	2.16
PM905Q02	Tennis balls	-0.46	0.60	0.30	0.26	-0.46	0.54	1.43

Note: List of items in which the difference of the solution rate in logits between the United States and the average of the reference countries is significantly beneath the average of these differences. The new items that did not appear in former comparisons with the OECD average (Table 3, Table 7 and Table 9) are in bold.

Source: Computed using the open source data from OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.

This method reveals five new conspicuous items, again in bold.

2.3 Differential item functioning (DIF) analysis

Here, DIF analysis is used in order to confirm the conspicuous items identified so far in a statistical way.

DIF can be defined as follows: in psychological or educational testing, abilities generally need to be measured by means of a specific test. The aim is to measure a particular ability for all the test takers in the same way. Therefore, one has to look for items that function differently for different groups. In contrast to simple mean differences for an item, "...DIF occurs when examinees from different groups show differing probabilities of success on (...) the item after matching on the underlying ability that the item is intended to measure." (Zumbo, 1999: 12). To put it into the words of item-response-theory (IRT), for two groups having the same ability (therefore the same mean of the individual person parameters) there should be no difference in the corresponding item parameters. Walker (2011: 365) pointed out that DIF arises "...when items on a test or psychological measure are multidimensional, measuring constructs or abilities or dimensions in addition to the primary dimension the assessment tool was designed to measure and two or more groups differ in their underlying distribution on these additional abilities."

In the literature there are several methods of detecting a DIF. For the approach used here, it was decided that an IRT-based method, which seems to be one of “the most promising of these methods (used for DIF detection, annotation of the authors)”, be used (Swaminathan and Rogers, 1990: 361).

The items were scaled in a 1PL-IRT-model.⁵ Theoretically, if an item has no DIF, the resulting item parameters for different groups should be the same (or at least vary only a little). This is exactly what the Waldtest (implemented in the package eRm for the programme R [Mair and Hatzinger, 2009]) does (Strobl, 2012; see also Fischer and Scheiblechner, 1970). In this test, the corresponding item parameters for the two groups are compared under control of their variance.⁶ The given output includes the calculated differences between the parameters (which follow the standard normal distribution) as well as the corresponding p-values indicating the significance. Positive values for the difference (following z-statistic) can be interpreted as relative strengths of the United States and conversely the negative ones as relative weaknesses. An overview with all PISA items and the corresponding statistics of the Waldtest would be too long for this report, therefore it is shown in the Annex (see Table 21).

Due to the huge sample size it was decided to work with a significance level of 1%.

Table 14. DIF-statistics for the items that are conspicuous in the logit approach

Item ID	Item name	z-statistic	p-value
PM998Q02	Bike rental	12.10656	0.00000
PM420Q01T	Transport	11.35140	0.00000
PM446Q01	Thermometer cricket	9.99092	0.00000
PM954Q01	Medicine doses	8.33662	0.00000
PM918Q01	Charts	8.31967	0.00000
PM411Q02	Diving	7.79967	0.00000
PM915Q01	Carbon tax	6.93497	0.00000
PM903Q03	Drip rate	6.78604	0.00000
PM955Q01	Migration	6.76779	0.00000
PM943Q01	Arches	6.68874	0.00000
PM155Q01	Population pyramids	6.68494	0.00000
PM909Q01	Speeding fines	6.50526	0.00000
PM828Q02	Carbon dioxide	6.38880	0.00000
PM155Q04T	Population pyramids	4.82974	0.00000
PM982Q02	Employment data	3.32424	0.00089
PM982Q01	Employment data	1.06484	0.28695
PM446Q02	Thermometer cricket	-0.91201	0.36176
PM00GQ01	An advertising column	-1.92636	0.05406
PM992Q03	Spacers	-2.35879	0.01833
PM955Q03	Migration	-3.87156	0.00011
PM949Q01T	Roof truss design	-4.42231	0.00001
PM995Q01	Revolving door	-4.81654	0.00000
PM943Q02	Arches	-5.17043	0.00000
PM924Q02	Sauce	-5.26597	0.00000

5. The R programme and the package eRm were used (Mair and Hatzinger, 2009; Strobl, 2012).

6. For a more detailed explanation, see Mair and Hatzinger (2009).

Table 14. DIF-statistics for the items that are conspicuous in the logit approach (*continued*)

Item ID	Item name	z-statistic	p-value
PM406Q01	Running tracks	-5.64003	0.00000
PM406Q02	Running tracks	-5.87354	0.00000
PM905Q02	Tennis balls	-6.60384	0.00000
PM305Q01	Map	-6.95330	0.00000
PM00KQ02	Wheelchair basketball	-7.33787	0.00000
PM464Q01T	The fence	-7.49525	0.00000
PM955Q02	Migration	-8.30818	0.00000
PM800Q01	Computer game	-8.36962	0.00000
PM915Q02	Carbon tax	-11.68626	0.00000

Source: Computed using the open source data from OECD (2012a), Cognitive item response data file, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/INT_COG12_DEC03.zip.

In Table 14, we only look at the 33 items that have been identified as conspicuous in the logit approach. Highlighted in grey are those items showing a significant DIF, which can be interpreted as the relative strengths of the United States. Conversely, the items marked in blue show the relative weaknesses of the United States. Four items are highlighted in italics; these items show no significant DIF at the 1% level.

The four items highlighted in italics have been conspicuous so far, but show no significant DIF at the 1% level. Nevertheless, three out of these four items can be justified by looking at DIFs referring to the five reference countries/economies. Items “Spacers 03” and “Thermometer cricket 02” especially, where the United States had a relative weakness, show quite large negative differences compared to Korea and Shanghai, China. This can be seen in Table 15.

Table 15. Items showing a large DIF in comparison to Asian countries/economies (negatively)

Item ID	Item name	Shanghai (China)		Korea	
		z-statistic	p-value	z-statistic	p-value
PM446Q02	Thermometer Cricket	-16.86502	0.00000	-4.55104	0.00001
PM992Q03	Spacers	-12.23709	0.00000	-16.54550	0.00000

Source: Computed using the open source data from OECD (2012a), Cognitive item response data file, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/INT_COG12_DEC03.zip.

The item “Employment data 01”, where the United States had relative strength, also shows quite a large difference to Korea and Shanghai, China, but this time positively for the United States. This can be seen in Table 16.

Table 16. Items showing a large DIF in comparison to Asian countries/economies (positively)

Item ID	Item name	Shanghai (China)		Korea	
		z-statistic	p-value	z-statistic	p-value
PM982Q01	Employment data	21.11924	0.00000	7.81034	0.00000

Source: Computed using the open source data from OECD (2012a), Cognitive item response data file, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/INT_COG12_DEC03.zip.

Only the item “An advertising column 01” cannot be justified as conspicuous via DIF-analysis. Nevertheless, this item will be used further in the analyses.

If all items with a DIF are examined (see Table 21 in the Annex) there are some items that have not been conspicuous so far and thus appear for the first time (see Table 17 and Table 18).

Table 17. Relative strengths appearing for the first time in DIF-analysis

Item ID	Item name	z-statistic	p-value
PM033Q01	A view room	2.70003	0.00693
PM155Q02D	Population pyramids	4.92041	0.00000
PM273Q01T	Pipelines	2.82496	0.00473
PM828Q01	Carbon dioxide	4,61569	0,00000
PM905Q01T	Tennis balls	4.60279	0.00000
PM906Q02	Crazy ants	5.30677	0.00000
PM918Q05	Charts	3.82277	0.00013
PM949Q03	Roof truss design	3,01219	0,00259
PM953Q02	Flu test	4.22350	0.00002
PM954Q02	Medicine doses	5.73154	0.00000

Source: Computed using the open source data from OECD (2012a), Cognitive item response data file, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/INT_COG12_DEC03.zip.

Table 18. Relative weaknesses appearing for the first time in DIF-analyses

Item ID	Item name	z-statistic	p-value
PM828Q03	Carbon dioxide	-3.18858	0.00143
PM903Q01	Drip rate	-4.85138	0.00000
PM906Q01	Crazy ants	-5.01058	0.00000
PM923Q01	Sailing ships	-2.93591	0.00333

Source: Computed using the open source data from OECD (2012a), Cognitive item response data file, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/INT_COG12_DEC03.zip.

Some of these items will be further used in the analyses for demarcating purposes in the patterns for relative strength or weaknesses, but not as items for the patterns themselves.

2.4 Summary of all conspicuous items

The following two tables show a summary of all 33 items from Table 14. The methods are listed in the columns:

1. diff.log_United States-OECD: Difference between the United States and the OECD average (logits)

diff.log_usa-avg_r.cnt: Difference between the United States and the average of the reference countries/economies (logits)

2. DIF- Analysis: The z-statistic shows the differences of the respective items for both groups (United States and all other participating countries/economies).
The p-value is given for the z-statistic which follows the standard normal distribution.

Table 19. List of conspicuous items - relative strengths

Relative strengths							
Item ID	Item name	solution rate United States	Avg. solution rate of all participating countries/economies	diff.log_United States-OECD	diff.log_usa- avg_r.cnt	DIF-analysis	
						z-statistic	p-value
PM155Q01	Population pyramids	69.39	67.78	0.08	-0.28	6.68494	0.00000
PM155Q04T	Population pyramids	53.92	55.88	N.S.	-0.31	4.82974	0.00000
PM411Q02	Diving	48.94	45.73	0.13	N.S.	7.79967	0.00000
PM420Q01T	Transport	57.82	50.03	0.31	0.23	11.35140	0.00000
PM446Q01	Thermometer cricket	74.74	68.59	0.30	-0.16	9.99092	0.00000
PM828Q02	Carbon dioxide	58.68	55.97	0.11	N.S.	6.38880	0.00000
PM903Q03	Drip rate	29.92	25.72	0.21	N.S.	6.78604	0.00000
PM909Q01	Speeding fines	92.91	89.35	0.45	-0.21	6.50526	0.00000
PM915Q01	Carbon tax	43.56	40.20	0.14	N.S.	6.93497	0.00000
PM918Q01	Charts	91.93	87.27	0.51	0.25	8.31967	0.00000
PM943Q01	Arches	51.17	50.04	N.S.	-0.32	6.68874	0.00000
PM954Q01	Medicine doses	70.33	65.42	0.23	N.S.	8.33662	0.00000
PM955Q01	Migration	73.54	72.13	N.S.	-0.24	6.76779	0.00000
PM982Q01	Employment data	87.23	87.30	N.S.	0.05	N.S.*	N.S.*
PM982Q02	Employment data	33.19	30.74	0.11	-0.09	3.32424	0.00089
PM998Q02	Bike rental	80.71	71.59	0.51	-0.23	12.10656	0.00000

Note: * The item "Employment data 01 does not show a significant "global" DIF at the 1% level, but shows a high DIF compared to Korea and Shanghai, China (see Table 16).

Source: Analysis done by the authors, summarising the findings of tables 14-18.

Table 20. List of conspicuous items - relative weaknesses

Relative weaknesses							
Item ID	Item name	solution rate_United States	Avg. solution rate of all participating countries/economies	diff.log_United States-OECD	diff.log_usa- avg_r.cnt	DIF-analysis	
						z-statistic	p-value
PM00GQ01	An advertising column	5.37	8.78	-0.53	N.S.	N.S.	N.S.
PM00KQ02	Wheelchair basketball	7.01	14.86	-0.84	-1.34	-7.33787	0.00000
PM305Q01	Map	46.45	60.37	-0.56	N.S.	-6.95330	0.00000
PM406Q01	Running tracks	16.69	25.63	-0.54	-1.14	-5.64003	0.00000
PM406Q02	Running tracks	8.87	16.90	-0.74	-1.37	-5.87354	0.00000
PM446Q02	Thermometer cricket	5.38	6.82	N.S.	-1.12	N.S.*	N.S.*
PM464Q01T	The fence	14.08	23.67	-0.64	-1.28	-7.49525	0.00000
PM800Q01	Computer game	79.49	88.41	-0.68	-1.12	-8.36962	0.00000
PM905Q02	Tennis balls	38.70	50.07	N.S.	-1.06	-6.60384	0.00000
PM915Q02	Carbon tax	52.94	68.26	-0.65	-1.27	-11.68626	0.00000
PM924Q02	Sauce	51.32	63.48	-0.50	N.S.	-5.26597	0.00000
PM943Q02	Arches	2.48	5.29	-0.79	-1.53	-5.17043	0.00000
PM949Q01T	Roof truss design	57.68	67.54	N.S.	-1.16	-4.42231	0.00001
PM955Q02	Migration	19.61	34.23	-0.76	-1.27	-8.30818	0.00000
PM955Q03	Migration	5.97	11.99	-0.76	-1.51	-3.87156	0.00011
PM992Q03	Spacers	5.43	8.11	N.S.	-1.13	N.S.*	N.S.*
PM995Q01	Revolving door	46.96	57.71	N.S.	-1.07	-4.81654	0.00000

Note: * The items "Employment data 01", "Thermometer cricket 02" and "Spacers 03" do not show a significant "global" DIF at the 1% level, but show a high DIF compared to Korea and Shanghai, China (see Table 15 and Table 16).

Source: Analysis done by the authors, summarising the findings of tables 14-18.

3. Patterns

The 33 conspicuous items identified in Section 2 (Table 19 and Table 20) will now be examined more closely. Do different items have the same or similar cognitive requirements, are there certain “*patterns*”? Of course, every single PISA item has its own characteristics. A pattern is, thus, composed of items that share some common features, whereas other features remain item specific. Seven such patterns will be identified, three for the relative strengths and four for the relative weaknesses of the United States. The big majority of *conspicuous items* (27 items) fit into one of these patterns. For each pattern, one item will be selected that is particularly “typical” for this pattern as an illustration; in the following, these items will be called “illustrating items”. For some of the patterns, a second item will be selected that is also fairly typical and belongs to the set of *released* PISA items so that it can be published here. The released illustrating items will be analysed in more detail in Section 4. In some of the patterns, we will also identify “demarcating items”. That means such an item seems, at a first glance, to belong to a certain pattern, but, in fact, requires different competencies in order to solve it. A *demarcating item* for a strength pattern is relatively harder for US students than the pattern suggests, and a *demarcating item* for a weakness pattern is relatively easier for US students.

There are three out of the 33 items that do not fit the patterns and therefore will no longer be used in this report: “Population pyramids 04”, “Diving 02”, and “Sauce 02”. Another three of the 33 items are used as “demarcating items” for some of the patterns: “Arches 01” for a weakness pattern, and “Computer game 01” and “Carbon tax 02” for strength patterns.

3.1 A: Relative strengths of the United States

It is remarkable that the following “strength patterns” include the majority of easy PISA items; there are only a few more challenging ones here. So, not only the absolute strengths (which is clear, because easy items are just items that are solved more often by the participating students), but also the relative strengths of the United States reveal themselves mostly in easy items. More precisely, the easiest PISA items are part of our first pattern A1. With this in mind, one might say that pattern A1 represents the strengths of the United States best. It includes, almost exclusively, easy items, in particular “Speeding fines 01”, “Charts 01”, “Employment data 01” and “Bike rental 02”, which have the highest solution rates for the United States of all (from 92.91% for “Speeding fines 01” to 80.71% for “Bike rental 02”).

Pattern A1: Read data directly from tables and diagrams (one step)

The items in this pattern require only understanding a short text and reading single values directly from a given representation such as a table or a bar diagram.

Items in this pattern	“Transport 01”, “Speeding fines 01”, “Charts 01”, “Migration 01”, “Bike rental 02” and “Employment data 01”
Illustrating items:	“Migration 01” and (released) “Charts 01”
Demarcating items:	“Computer game 01” and “Running time 01”

There are two more items that, at first glance, seem to fit this pattern as well, but which are conspicuously negative when comparing the United States with the OECD average: “Computer game 01” and “Running time 01”. A brief description will be given as to why these items do not fit this pattern. “Computer game 01” not only represents a weakness of the United States but, in fact, shows a significant negative difference. In order to get this item correct, more than one step is needed and the first step is not reading data directly. Rather, it is first necessary to calculate total scores and then compare these in a second step. Similarly, in “Running time 01”, more than one step is needed to solve this item. The numbers must first be arranged according to their size and then the third lowest number has to be identified.

Pattern A2: Simple handling of data from tables and diagrams

The items in this pattern require understanding a short text, reading two values from a given representation and then performing some straightforward operation such as adding or comparing the values.

Items in this pattern	"Population pyramids 01", "Carbon dioxide 02", "Carbon tax 01", and "Employment data 02"
Illustrating item:	"Carbon tax 01"
Demarcating items:	"Charts 02", "Charts 05" and "Carbon dioxide 03"

There are also some items for this pattern that seem to fit at first sight: "Charts 02", "Charts 05" and "Carbon Dioxide 03". However, the first two items require more than a straightforward operation. In "Charts 02", a whole development process must be considered over time and "Charts 05" requires the continuation of a trend. In addition, "Carbon dioxide 03" is more challenging, since it is necessary here to handle units (the data in the diagram are given in millions of metric tons) where it is easily possible to make mistakes.

Pattern A3: Handling directly manageable formulae

The items in this pattern require using a given formula, e.g. inserting numbers for variables and doing some easy calculations. The formula can be used directly, without any re-structuring.

Items in this pattern	"Thermometer cricket 01", "Drip rate 03" and "Medicine doses 01"
Illustrating items:	"Medicine doses 01" and (released) "Drip rate 03"
Demarcating items:	"Carbon tax 02", "Medicine doses 04" and "Drip rate 01"

The items "Carbon tax 02", "Medicine doses 04" and "Drip rate 01" seem to fit into this pattern as well, since there are formulae involved. However, on closer inspection they do not. In "Carbon tax 02", the given formula is not directly manageable because it is necessary to apply some arithmetic laws first ("x before +"). Ignoring these will easily lead to mistakes. In "Medicine doses 04", there are two formulae that have to be compared, the values have to be switched from years to months for the second formula, and the values have to be rounded off appropriately. The item "Drip rate 01" contains a formula that has to be interpreted functionally ("what happens if..."), so it does not fit into this pattern either.

3.2 B: Relative weaknesses of the United States

It is remarkable that nearly all of the most difficult PISA items are part of the patterns that display not only absolute weaknesses (which is clear, because difficult items are just items that are solved less often by the participating students), but the relative weaknesses of the United States. The five items with the lowest solution rates according to the OECD average are "Revolving door 02", "Arches 02", "Thermometer cricket 02", "Spacers 03", and "An advertising column" (from 3.5% for "Revolving door 02" up to 8.8% for "An advertising column"). All these items turn up in the following patterns. This leads to the presumption that the United States has a particular weakness in challenging items.

Pattern B1: Curricular requirement: π is necessary

The items in this pattern require the explicit use of the circle number π in a calculation.

Items in this pattern	"Wheelchair basketball 02", "Running tracks 01", "Running tracks 02", "An advertising column 01" and "Revolving door 02" (part of this pattern, but slightly not significant)
Illustrating item:	"Running tracks 01" and (released) "Revolving door 02"
Demarcating item	"Arches 01"

To underline that this pattern is the result of a curricular requirement (being familiar with the circle number π) we chose “Arches 01” as a demarcating item. As in all other items in this pattern it is necessary to identify the concrete geometric form of a circle and to operate with it in some way. In the item “Arches 01”, a circular arch is given and one has to find the relation of the height (radius) and the width (diameter). Therefore π is not necessary here, although π is used in two of the four possible answers in this multiple-choice item. So circles themselves seem not to be the problem for the US students, since they did particularly well in “Arches 01” (see Table 19).

Pattern B2: Substantial mathematisation of a real world situation

The items in this pattern require establishing a mathematical model of a given real world situation in the form of a term or an equation with variables for (geometric or physical) quantities, before further actions (especially calculations) can take place. The situation has to be understood and the appropriate mathematical content has to be activated and applied.

Items in this pattern	“Thermometer cricket 02”, “The fence 01”, “Arches 02”, “Spacers 03” and “Revolving door 02”
Illustrating item	“Arches 02” and (released) “Revolving door 02”

Pattern B3: Genuine interpretation of real world aspects

The items in this pattern require taking a given real world situation seriously and genuinely interpreting aspects of it. The popular superficial classroom strategy “Don’t care about the context, just extract the numbers from the text and do some obvious operations” fails here.

Items in this pattern	“Tennis balls 02”, “Migration 02”, “Migration 03”, (“An advertising column 01”)*, (“Running Tracks 02”) and (“Drip rate 01”)
Illustrating item	“Migration 02”

Note: *Second category items are marked with brackets and primarily fit in another pattern.

Pattern B4: Reasoning in a geometric context

The items in this pattern require genuine reasoning in a planar or spatial geometric context by using geometric concepts and facts. The transformation from the given real world context into the corresponding geometric context is straightforward and not the hard part of the task.

Items in this pattern	“Map 01”, “Roof truss design 01”, “Revolving door 01”, (“Arches 02”), (“Running tracks 01”) and (“Running tracks 02”)
Illustrating item	“Map 01” and (released) “Revolving door 01”

4. Analyses of illustrating items

In this section, we will illustrate some of the patterns identified in Section 3 (those for which released items are available) by typical items. Thus, the patterns ought to become more transparent and more concrete. As was already said in Section 3, each single item in the PISA test has its own structure and its specific cognitive requirements. A certain pattern is, therefore, composed of items that share some common features, whereas other features are item specific. The illustrating items contain these common features, besides other features that are not typical for the pattern.

4.1 Strength patterns (A1 and A3)

Pattern A1: Read data directly from tables and diagrams (one step)

The items in this pattern require only understanding a short text and reading single values directly from a given representation, such as a table or a bar diagram.

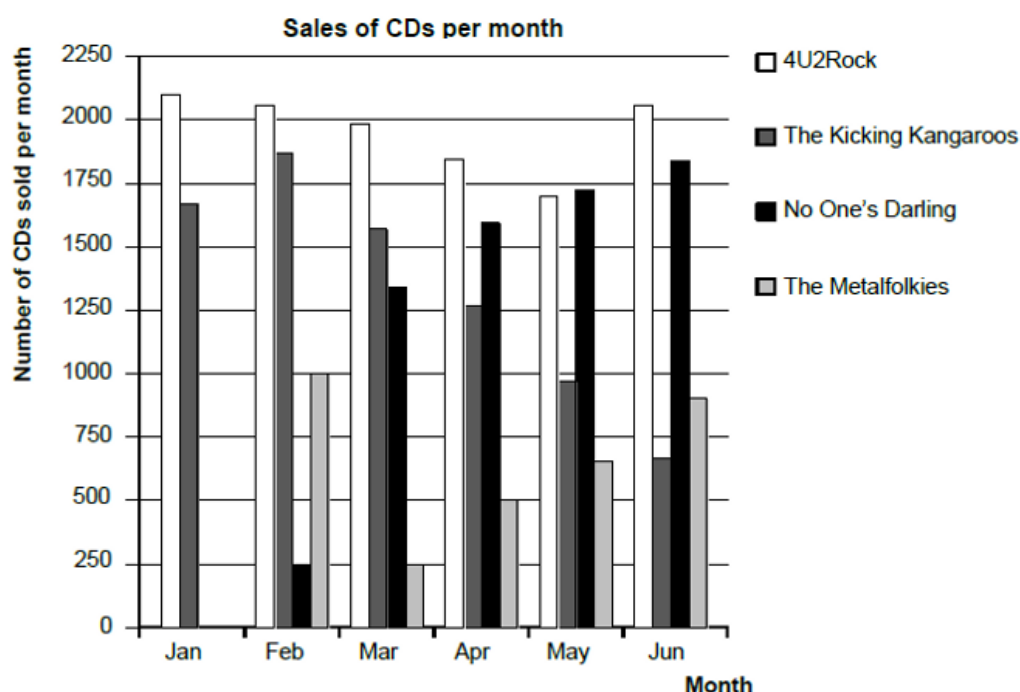
To illustrate this pattern we have chosen the item “Charts 01”. With its solution rates of 91.93% for the United States and 87.27% for the OECD average, it is a rather easy item, but the items “Employment

data 01” and “Speeding fines 01” in this pattern have even higher solution rates. Only “Transport 01”, with 57.82% for the United States and 50.03% for the OECD average, has a much lower solution rate (because of its complex-multiple-choice structure – the single statements in “Transport 01” are all easy). “Charts 01” requires interpreting given data and belongs to the mathematical content area “Uncertainty and data”. This “Content” and “Process” combination represents a specific strength of the United States students (see Subsection 2.2.2).

Figure 4. Charts (PM918Q01)

CHARTS

In January, the new CDs of the bands *4U2Rock* and *The Kicking Kangaroos* were released. In February, the CDs of the bands *No One's Darling* and *The Metalfolkies* followed. The following graph shows the sales of the bands' CDs from January to June.



Question 1: CHARTS

PM918Q01

How many CDs did the band *The Metalfolkies* sell in April?

- A 250
- B 500
- C 1000
- D 1270

Source: OECD (2013b), “PISA 2012 Released Mathematics Items”, May 2013, www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf.

Ideal-typical solution

In order to answer the question one needs to have a look at the April bars and to identify the fourth bar as representing the number of CDs sold by The Metalfolkies.

This bar reaches up to 500; therefore, the correct answer is: B

The task “Charts 01” is embedded in a real world situation and starts with a short stimulus that explains the structure of the following diagram. The number of CDs sold per month from four different bands is given for the six months from January to June. For each month from February on, there are four bars. Each bar represents the number of CDs sold in a specific month by a single band. The task for the students in item 01 is to find out how many CDs the band The Metalfolkies sold in April. Therefore, one needs to have a look at the bright grey bar for April, which represents the band The Metalfolkies. It reaches up to 500 so answer B is correct.

Pattern A3: Handling directly manageable formulae

The items in this pattern require using a given formula, e.g. inserting numbers for variables and doing some easy calculation. The formula can be used directly, without any re-structuring.

Here, the item “Drip rate 03” will be analysed to exemplify this pattern. With 29.92%, this item has a relatively low solution rate for the United States. However, with 25.72% for the OECD average, the solution rate is, overall, relatively low. The items in this pattern are exclusively from the mathematical content area “Change and relationships”, and the process of “Drip rate 03” is “Employ”. This combination represents another specific strength of the US students (see Subsection 2.2.2).

Figure 5. Drip rate (PM903Q03)

DRIP RATE

Infusions (or intravenous drips) are used to deliver fluids and drugs to patients.



Nurses need to calculate the drip rate, D , in drops per minute for infusions.

They use the formula $D = \frac{dv}{60n}$ where

d is the drop factor measured in drops per millilitre (mL)

v is the volume in mL of the infusion

n is the number of hours the infusion is required to run.

Question 3: DRIP RATE

PM903Q03 – 0 1 9

Nurses also need to calculate the volume of the infusion, v , from the drip rate, D .

An infusion with a drip rate of 50 drops per minute has to be given to a patient for 3 hours. For this infusion the drop factor is 25 drops per millilitre.

What is the volume in mL of the infusion?

Volume of the infusion: mL

Source: OECD (2013b), "PISA 2012 Released Mathematics Items", May 2013, www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf.

Ideal-typical solution

First one has to identify the given values and to find out which value is wanted. Then the formula can be converted for v so that afterwards one has only to insert the right values in the new formula and calculate the result. For this task, D is 50 [drops per minute], d is 25 [drops per mL] and n is 3 [hours]. Thus, $D = \frac{dv}{60n}$, $v = \frac{D \times 60n}{d} = \frac{50 \times 60 \times 3}{25} = 360$ [mL].

Therefore, the correct answer is: 360

The stimulus of the task "Drip rate" is used as an introduction to the topic and an explanation of a formula used by nurses to calculate the drip rate (in drops per minute) for infusions. This formula contains

four variables that are each separately explained in the stimulus. One of this item's challenges is to identify the relevant variables and their values. Another requirement is to insert these values correctly into the formula and then calculate the unknown volume of the infusion. This can either be achieved by inserting all given values and then converting the formula, or by converting the formula first and inserting the given values afterwards. This rearranging of the formula is another challenge of this item. The insertion is straightforward and the calculation is easy to perform. The formula is directly manageable since it is pre-structured by the fraction line. However, the relatively low solution rates in comparison to other items in this pattern are due to the number of necessary algebraic steps.

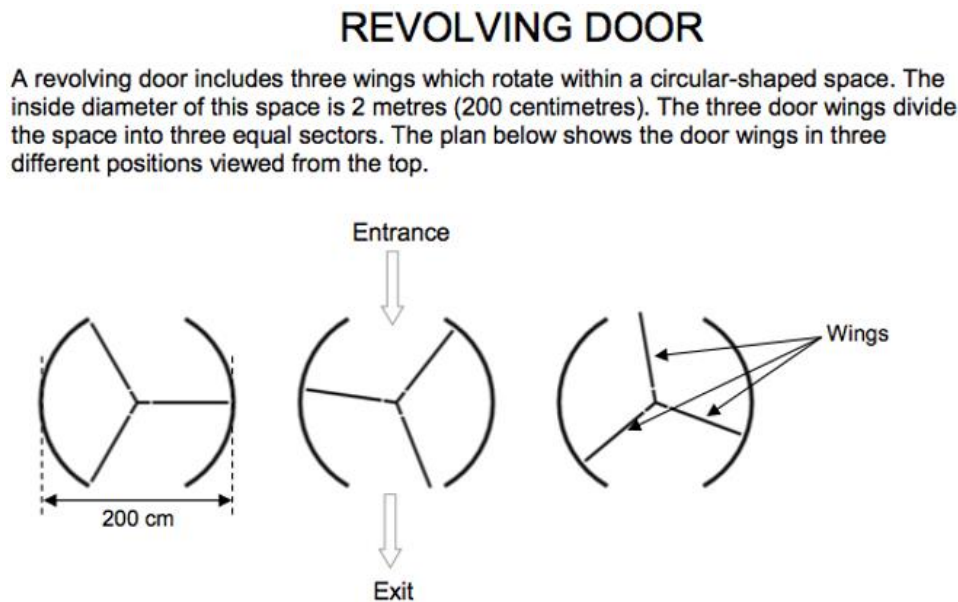
4.2 Weakness patterns (B1, B2 and B4)

In order to illustrate the patterns for the relative weaknesses of the United States, we will analyse the items 01 and 02 of the task "Revolving door". Item 02 has a wide range of cognitive requirements and therefore exemplifies several patterns. In the following, an explanation will be given regarding which of the typical features of the patterns are part of the items "Revolving door 01" and "Revolving door 02".

Analysis of the item "Revolving door 01"

This item will exemplify pattern B4 (Reasoning in a geometric context). The United States has a solution rate of 46.96%, whereas the OECD average is 57.71%. Therefore, this is an item with an intermediate level of difficulty. In pattern B4, three more difficult items ("Arches 02", "Running tracks 01", "Running tracks 02") can be found, in addition to an easier item ("Roof truss design 01") and an item with similar difficulty ("Map 01"). The mathematical content area of the item "Revolving door 01" is "Space and shape" – like all of the items in this pattern – and the process is "Employ".

Figure 6. Revolving door 01 (PM995Q01)



Question 1: REVOLVING DOOR

PM995Q01 – 0 1 9

What is the size in degrees of the angle formed by two door wings?

Size of the angle:°

Source: OECD (2013b), "PISA 2012 Released Mathematics Items", May 2013, www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf.

Ideal-typical solution

The door wings divide the circular-shaped space into three equal sectors; consequently, the angle formed by two door wings is $\frac{360^\circ}{3} = 120^\circ$.

Therefore, the correct answer is: 120° .

The stimulus of the item describes a revolving door and presents three diagrams that include information such as the diameter and different positions of the door wings. This description already refers to a mathematical model that is the geometric form of a circle, which is divided into three congruent parts by three linear segments. One difficulty is the rotation of the door wings, meaning there is no fixed shape and the students are faced with three slightly different diagrams instead of one. Even if the students do not have to mentally revolve the door wings, they need to recognise that the angle between the door wings stays the same when the door is rotated. Such a mental operation with the moving door wings is central to the second item of this task “Revolving door 02” (see below).

The essential cognitive requirement for “Revolving door 01” is to detect that the three sectors of the revolving door are congruent and, therefore, have the same angles. After that, the students have to conclude that one of the angles will be 120° . For this argument, the students need a knowledge of angles, namely that a full circle is 360° . A further difficulty could be that it might be unfamiliar for students to determine angles in sectors of a circle. These requirements are mainly features of pattern B4 (Reasoning in a geometric context).

Some mistakes can be expected. For example, that some students do not understand the geometric model and therefore simply pick out the only given value, which is the diameter (expected wrong solution 200°). Some students will recognise that they have to divide the full angle by 3 but do not know how many degrees in a full angle or do not have a proper concept of angles, so they might use an avoidance strategy, such as simply using the value of the diameter for the full angle (expected wrong solution 66.6°). Another improper strategy would be to guess or to measure the angle.

Analysis of the item “Revolving door 02”

In the following, the item “Revolving door 02” will be analysed in order to exemplify, in particular, patterns B1 (Curricular requirement: π is necessary) and B2 (Substantial mathematisation of a real world situation). “Revolving door 02” has the lowest solution rate for the OECD average in the whole test. The solution rate of the United States for this item is only 2.26% and the solution rate of the OECD average is 3.47%. The mathematical content area is “Space and shape” and the process is “Formulate”. This content area represents a particular weakness of the United States (see Subsection 2.2.2).

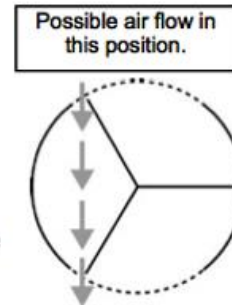
Figure 7. Revolving door 02 (PM995Q02)

Question 2: REVOLVING DOOR

PM995Q02 – 0 1 9

The two door **openings** (the dotted arcs in the diagram) are the same size. If these openings are too wide the revolving wings cannot provide a sealed space and air could then flow freely between the entrance and the exit, causing unwanted heat loss or gain. This is shown in the diagram opposite.

What is the maximum arc length in centimetres (cm) that each door opening can have, so that air never flows freely between the entrance and the exit?

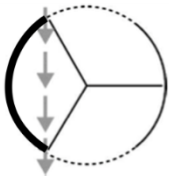


Maximum arc length: cm

Source: OECD (2013b), "PISA 2012 Released Mathematics Items", May 2013, www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf.

Ideal-typical solution

The minimal arc length on the left side of the door has to be one third of the circumference because the sector formed by two door wings is – according to the conditions of air flow – at least one third of the whole circle.



For reasons of symmetry, the minimal arc length on the right side of the door is also one third of the circumference.



Thus, at most, one third of the circumference is left for the two openings, which implies that the maximum opening for exit and entrance is one sixth of the circumference each:

$$\text{Circumference} = 200 \times \pi \approx 628 \text{ cm} \quad \text{Arc length} \approx 628.32 \div 6 \approx 104 \text{ cm}$$

Therefore, the correct answer is (rounded off): 104 cm.

The stimulus and the first item of this task have been analysed simply as an illustration for pattern B4.

Like “Revolving door 01”, the item “Revolving door 02” describes a real world situation by a text and a diagram. Neither the text nor the diagram in “Revolving door 02” contain the necessary value of the diameter, so the students have to transfer this information from the stimulus to the new situation in this item. The diagram leads to a mathematic model, but a substantial difficulty is that the item requires a

dynamic conception, because the lengths of the openings are flexible and the rotation of the door wings has to be considered. In order to identify, in particular, the arc length of the right side of the circle, it is necessary to modify the diagram and transfer it to a different position of the door wings. Therefore the mathematisation of the real world situation is much more difficult than in the first part of the task “Revolving door 01”. These cognitive requirements are typical for items of pattern B2 (Substantial mathematisation of a real world situation). For that reason, the item “Revolving door 02” can be used to exemplify this pattern.

It can be expected that a lot of the students are not able to formulate an adequate mathematic model of the given situation. Some students may use an improper model, e.g. to calculate the length of the arc between two door wings instead of the length of the arc of each opening. Furthermore, the task is unfamiliar, with no approaches or instructions given, so the students have to invent their own strategies. For example, it is possible to consider two extreme cases with the door at the entrance nearly opened and just closed; another possibility is to use a “dividing strategy”, that is, to divide the circle into six equal parts and to argue that entrance and exit have to be one sixth of the circumference each.

No matter which strategy is used, it is necessary in order to solve the item to calculate the circumference of the circle. For this calculation, the students need to know how to use the formula of the circumference (π times diameter or 2π times radius). This requirement is the core of pattern B1 (Curricular requirement: π is necessary) and therefore the item “Revolving door 02” is also suitable to exemplify pattern B1.

It can be expected that students who are not able to use the formula of the circumference will frequently use empirical strategies to measure the length. For example, they may try to measure the length with a ruler, or may use the given diameter as a scale.

5. Analysis of student solutions

In this section we will analyse a random sample of 500 student solutions for two illustrating items that include important features of the B-patterns in order to make the results reported so far more transparent.

5.1 “Revolving door 01”

A look at the student solutions of “Revolving door 01” reveals three mistakes that occur quite often. The reasons behind most of the wrong answers cannot be identified because most students have only written their answer without indicating a reason for it (it was not required in this item to show the complete calculation).

One of the most common (visible) mistakes is the answer “240”. The students divided the circle correctly into three parts of 120° each and then multiplied one segment by two. The question states “What is the size in degrees of the angle formed by two door wings?”. It is possible that these students misunderstood the word “wings” and thus calculated the angle of two segments instead of the angle between two wings. It is also possible that they simply regarded the “two” in the question as a prompt to multiply their first result 120° by 2 without looking at the geometric situation.

This mistake is shown in the two following solutions and exemplifies pattern B4 (Reasoning in a geometric context).

Figure 8. Revolving door 01 solution

REVOLVING DOOR

A revolving door includes three wings which rotate within a circular space. The inside diameter of this space is 2 meters (200 centimeters). The three door wings divide the space into three equal sectors. The plan below shows the door wings in three different positions viewed from the top.

200 cm

Entrance

Exit

Wings

Whole = 360
each one = 120
two together = 240

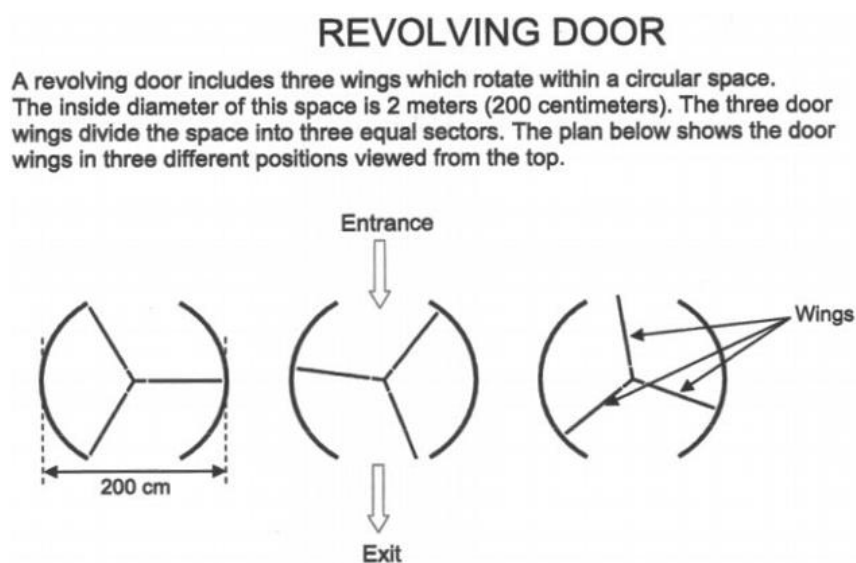
Question 38: REVOLVING DOOR PM995Q01 – 0 1 9

What is the size in degrees of the angle formed by two door wings?

Size of the angle: 240.....°

OECD (2013c), "Strengths and Weaknesses of American Students in Mathematics", in *Lessons from PISA 2012 for the United States*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264207585-5-en>, p. 69.

Figure 9. Revolving door 01 solution

**Question 9: REVOLVING DOOR**

PM995Q01 – 0 1 9

What is the size in degrees of the angle formed by two door wings?

Size of the angle: 240°

$$360^\circ \div 3 = 120^\circ$$

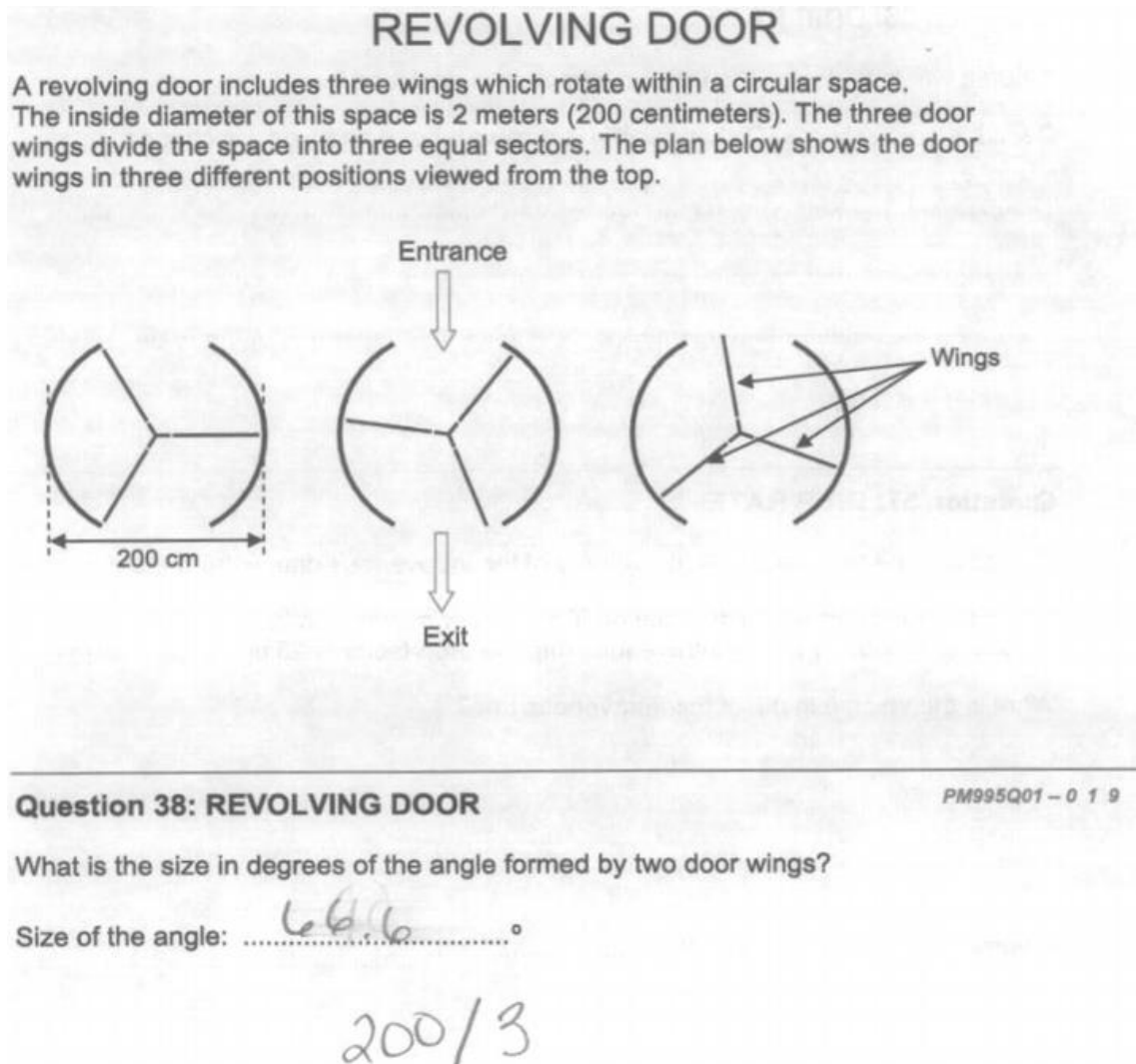


OECD (2013c), "Strengths and Weaknesses of American Students in Mathematics", in *Lessons from PISA 2012 for the United States*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264207585-5-en>, p. 69.

Another common mistake is "100". It can be assumed that these students either think that a circle has 300° in total, or they were simply using the length of the radius as a value for the angle, meaning that, instead of applying reasoning in this geometric situation, they simply took the given diameter (200) and divided it by 2 (because of the "two door wings").

The answers "66.7, 66.6 and 67" represent a third common mistake. For these answers it can be assumed that the students superficially divided 200 by 3, because 200 cm and three (doors) are the numbers given in the text. These students did not reason in the geometric context, but simply extracted the number 200 from the stimulus and did the obvious operation of dividing it by three. One such mistake is shown in the following solution.

Figure 10. Revolving door 01 solution



OECD (2013c), "Strengths and Weaknesses of American Students in Mathematics", in *Lessons from PISA 2012 for the United States*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264207585-5-en>, p. 70.

5.2 "Revolving door 02"

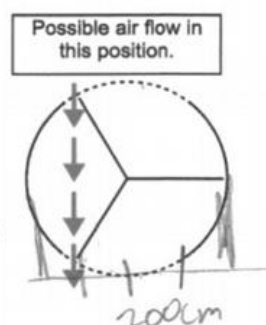
Our sample of 500 solutions contains a great number of varied answers and also many missing answers, which is certainly due to the high difficulty of the item (also, in this item it was not required that the students show their work). The most common answer was 100 cm, which is the radius of the circle. Since the diameter is the only given value it can be expected that a lot of students used it to estimate the arc length of one of the openings. An example is the following solution.

Figure 11. Revolving door 02 solution

Question 24: REVOLVING DOOR

The two door **openings** (the dotted arcs in the diagram) are the same size. If these openings are too wide the revolving wings cannot provide a sealed space and air could then flow freely between the entrance and the exit, causing unwanted heat loss or gain. This is shown in the diagram opposite.

What is the maximum arc length in centimeters (cm) that each door opening can have, so that air never flows freely between the entrance and the exit?



Maximum arc length: 100..... cm

OECD (2013c), "Strengths and Weaknesses of American Students in Mathematics", in *Lessons from PISA 2012 for the United States*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264207585-5-en>, p. 71.

In this solution, the student took the diameter and measured the approximate arc length of the opening. This solution illustrates that the student reconstructed, but over-simplified the situation. He/she tried to figure out the arc length of the opening, but ignored the curvature, or replaced it by a linear segment. Apparently he/she is not able to calculate the circumference.

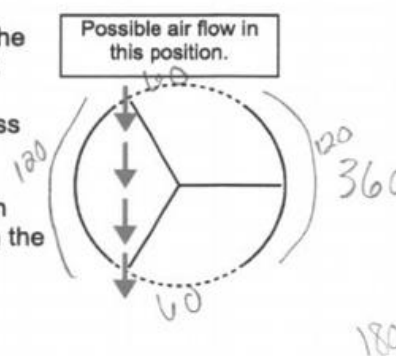
The next student solution also shows that the student was not able to deal with arc lengths. Indeed, the solution displays an adequate mathematical model of the situation, and 60° is the correct angle, but the student simply used angles instead of arc lengths.

Figure 12. Revolving door 02 solution

Question 24: REVOLVING DOOR

The two door **openings** (the dotted arcs in the diagram) are the same size. If these openings are too wide the revolving wings cannot provide a sealed space and air could then flow freely between the entrance and the exit, causing unwanted heat loss or gain. This is shown in the diagram opposite.

What is the maximum arc length in centimeters (cm) that each door opening can have, so that air never flows freely between the entrance and the exit?



Maximum arc length: 60..... cm

OECD (2013c), "Strengths and Weaknesses of American Students in Mathematics", in *Lessons from PISA 2012 for the United States*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264207585-5-en>, p. 72.

Both solutions, therefore, exemplify the requirements of pattern B1 (Curricular requirement: π is necessary). The fact that 100 cm and 60 cm were very often mentioned leads to the assumption that a lot of US students were not able to calculate the circumference. Of course, since most of the students only wrote down the result, it is difficult to identify the underlying mistakes.

The second most common answer was 200 cm, which is the value of the diameter. The students with this solution just indicated the value 200 cm without further explanations or considerations written down. They most likely did not take the situation seriously and just picked out the only given value.

Figure 13. Revolving door 02 solution

Maximum arc length: 200..... cm

OECD (2013c), "Strengths and Weaknesses of American Students in Mathematics", in *Lessons from PISA 2012 for the United States*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264207585-5-en>, p. 72.

This might be caused by the superficial classroom strategy "Don't care about the context, just extract the numbers from the text and do some calculation"; therefore this solution also fits pattern B3 (Genuine interpretation of real world aspects).

Nevertheless, there are also solutions that display a correct calculation of the circumference. The following two solutions exemplify that some of the US students have been able to calculate the circumference, but had problems understanding the real world situation, or had not been able to formulate an adequate mathematical model. So these two solutions do not fit into pattern B1 (Curricular requirement: π is necessary), but into pattern B2 (Substantial mathematisation of a real world problem).

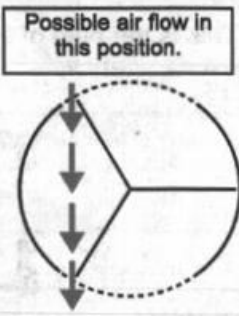
Figure 14. Revolving door 02 solution

Question 24: REVOLVING DOOR PM995Q02 - 0 1 9

The two door openings (the dotted arcs in the diagram) are the same size. If these openings are too wide the revolving wings cannot provide a sealed space and air could then flow freely between the entrance and the exit, causing unwanted heat loss or gain. This is shown in the diagram opposite.

What is the maximum arc length in centimeters (cm) that each door opening can have, so that air never flows freely between the entrance and the exit?

Possible air flow in this position.



Handwritten solutions:

$$C = 2\pi r$$

$$C = 2\pi 100$$

$$C = 200\pi$$

$$200\pi/3 = 209.4395$$

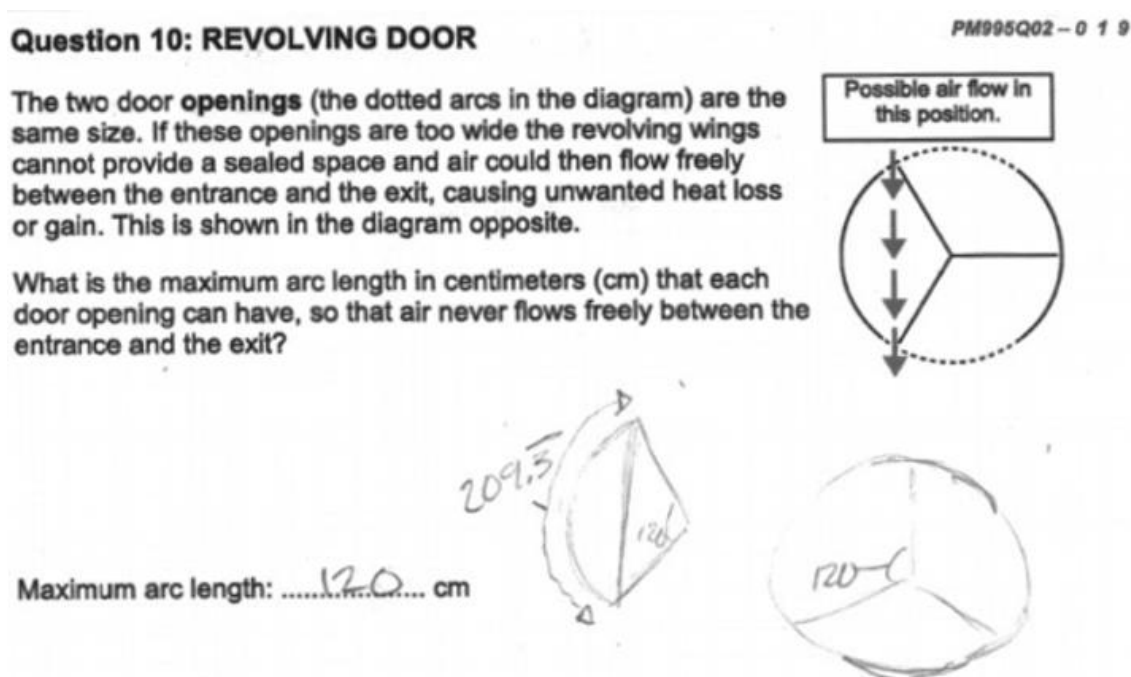
Maximum arc length: 209..... cm

OECD (2013c), "Strengths and Weaknesses of American Students in Mathematics", in *Lessons from PISA 2012 for the United States*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264207585-5-en>, p. 73.

The answer 209 cm was quite frequent. This student first calculated the circumference and then divided it by 3. The result is the arc length of one sector. Maybe the student did not understand the situation and therefore used an improper mathematic model. But it is also possible that he/she recognised that the length of the opening of the entrance and exit in total is one third of the circumference.

Slightly different is the next student solution.

Figure 15. Revolving door 02 solution



OECD (2013c), "Strengths and Weaknesses of American Students in Mathematics", in *Lessons from PISA 2012 for the United States*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264207585-5-en>, p. 73.

Like the solution above, the student calculated one third of the circumference, which is about 209 cm. In contrast to the former solution, it is clearly visible that the student calculated the arc length of a sector formed by two door wings and not the total length of the openings. It is noticeable that the student then wrote down the angle of the sector instead of the arc length, so this solution reveals two different student problems. The first problem is the usage of an improper mathematic model, namely to consider the arc length of a sector instead of an opening. The second problem seems to be confusion between angles and arcs.

6. Summary

The aim of this report was to identify relative strengths and weaknesses of US students in the PISA 2012 mathematics test. For that purpose, we have compared the US solution rates (technically transferred to logits in order to get a linear scale) for all 84 PISA mathematics items that have been administered in the United States with the solution rates of the OECD average and of five reference countries/economies that have performed significantly above the OECD average and are, economically speaking, important partners and also competitors of the United States: Canada; Germany; Korea; the Netherlands; and Shanghai, China. We have found altogether 33 "conspicuous" items where the differences between the United States and the OECD or the reference countries/economies are in some way statistically significant, and thus

identified certain patterns of relative strengths and relative weaknesses, where each pattern contains items with similar cognitive requirements. For 16 of these 33 items, the United States has performed relatively well; this has resulted in three strength patterns. For 17 items, the United States has performed relatively badly; this has resulted in four weakness patterns.

It is, therefore, interesting to see that the US students do not perform uniformly compared with the OECD average or with those reference countries/economies. Altogether, the US average (43.8%) in PISA mathematics is below the OECD average (47.5%) and further below the averages of the other five countries/economies. However, there are some items where the US students have performed nearly as well as Shanghai, China and better than the OECD average, as well as better than two, three or even four of the other reference countries/economies; these items constitute the relative strengths of the United States. On the other hand, there are items where the United States is much further away from the OECD average and from the reference countries/economies than the average differences may suggest; these items constitute the relative weaknesses of the United States.

It is remarkable that the strength patterns consist nearly exclusively of easy PISA items and the weakness patterns mostly of demanding items. Both the absolutely easiest and hardest PISA items occur in our patterns, the easiest in the strength and the hardest in the weakness patterns. This is not at all self-evident. It is clear that the United States, like all other countries/economies, performs rather well in easy items and rather moderately in difficult items; this is the definition of “easy” and “difficult”. However, it is not at all clear that the United States performs relatively well or badly in these items. It seems that the US students have particular strengths in cognitively less demanding mathematical skills and abilities, such as extracting single values from diagrams or handling well-structured formulae. And they have particular weaknesses in demanding skills and abilities, such as taking real world situations seriously, transferring them into mathematical terms and interpreting mathematical aspects in real world problems. These are tasks where the well-known superficial classroom strategy “Don’t care about the context, just extract the numbers from the text and do some obvious operations” is definitely bound to fail. This strategy is popular all over the world and frequently helps pupils and students to survive in school mathematics and to pass examinations. However, in a typical PISA mathematical literacy task, the students have to use the mathematics they have learned in a well-founded manner. The US students obviously have particular problems with such tasks. Of course, in more demanding tasks, some basic skills, such as the ones mentioned (extracting values or handling formulae) are needed, too, so the relative strengths of the US students are a necessary prerequisite for solving higher order tasks. Therefore, when it comes to implications of these findings, one clear recommendation would be to focus much more on higher order activities, such as those involved in mathematical modelling (understanding real world situations, transferring them into mathematical models, and interpreting mathematical results), without neglecting the basic skills needed for these activities.

REFERENCES

- Fischer, G. H. and H. Scheiblechner (1970), “ Algorithmen und Programme für das probabilistische Testmodell von Rasch”, *Psychologische Beiträge*, Vol. 12, pp. 23-51.
- Freudenthal, H. (1983), *Didactical Phenomenology of Mathematical Structures*, Reidel, Dordrecht.
- Mair, P. and R. Hatzinger, R (2009), “Extended Rasch modeling: The R package eRm”, <http://cygwin.uib.no/pub/cran/web/packages/eRm/vignettes/eRm.pdf>.
- Niss, M. and T. Højgaard (eds.) (2011), “Competencies and mathematical learning: Ideas and inspiration for the development of mathematics teaching and learning in Denmark”, *IMFUF Atekst*, No. 485-2011, Roskilde University, Roskilde.
- OECD (2013a), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264190511-en>.
- OECD (2013b), “PISA 2012 Released Mathematics Items”, May 2013, www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf.
- OECD (2013c), “Strengths and Weaknesses of American Students in Mathematics”, in *Lessons from PISA 2012 for the United States*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264207585-5-en>.
- OECD (2012a), Cognitive item response data file, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/INT_COG12_DEC03.zip.
- OECD (2012b), Compendium for the cognitive item responses database, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/M_comp_COG_DEC03.zip.
- Strobl, C. (2012), *Das Rasch-Modell: Eine verständliche Einführung für Studium und Praxis*, Rainer Hampp Verlag, München.
- Swaminathan, H. and H.T. Rogers (1990), “Detecting differential item functioning using logistic regression procedures”, *Journal of Educational Measurement*, Vol. 27(4), pp. 361-370.
- Walker, C. M. (2011), “What’s the DIF? Why differential item functioning analyses are an important part of instrument development and validation”, *Journal of Psychoeducational Assessment*, Vol. 29(4), pp. 364-376.
- Zumbo, B. D. (1999), *A Handbook on the Theory and Methods of Differential Item Functioning (DIF)*, National Defense Headquarters, Ottawa.

ANNEX: DIF STATISTICS

Table 21. Overview of the Waldtest statistics for DIF-detection sorted by size

Item code	Item name	z-statistic	p-value
PM998Q02	Bike rental	12.10656	0.00000
PM420Q01T	Transport	11.35140	0.00000
PM446Q01	Thermometer cricket	9.99092	0.00000
PM954Q01	Medicine doses	8.33662	0.00000
PM918Q01	Charts	8.31967	0.00000
PM411Q02	Diving	7.79967	0.00000
PM915Q01	Carbon tax	6.93497	0.00000
PM903Q03	Drip rate	6.78604	0.00000
PM955Q01	Migration	6.76779	0.00000
PM943Q01	Arches	6.68874	0.00000
PM155Q01	Population pyramids	6.68494	0.00000
PM909Q01	Speeding fines	6.50526	0.00000
PM828Q02	Carbon dioxide	6.38880	0.00000
PM954Q02	Medicine doses	5.73154	0.00000
PM906Q02	Crazy ants	5.30677	0.00000
PM155Q02D	Population pyramids	4.92041	0.00000
PM155Q04T	Population pyramids	4.82974	0.00000
PM828Q01	Carbon dioxide	4.61569	0.00000
PM905Q01T	Tennis balls	4.60279	0.00000
PM953Q02	Flu test	4.22350	0.00002
PM918Q05	Charts	3.82277	0.00013
PM982Q02	Employment data	3.32424	0.00089
PM949Q03	Roof truss design	3.01219	0.00259
PM273Q01T	Pipelines	2.82496	0.00473
PM033Q01	A view room	2.70003	0.00693
PM995Q03	Revolving door	2.56107	0.01044
PM496Q01T	Cash withdrawal	2.53601	0.01121
PM571Q01	Stop the car	2.50056	0.01240
PM803Q01T	Labels	2.34191	0.01919
PM918Q02	Charts	2.33155	0.01972
PM442Q02	Braille	1.62494	0.10418
PM462Q01D	Third side	1.51383	0.13007
PM919Q02	Z's fan merchandise	1.28567	0.19856
PM564Q01	Chair lift	1.16281	0.24490

Table 21. Overview of the Waldtest statistics for DIF-detection sorted by size (continued)

Item code	Item name	z-statistic	p-value
PM949Q02T	Roof truss design	1.09568	0.27322
PM982Q01	Employment data	1.06484	0.28695
PM411Q01	Diving	0.93519	0.34969
PM954Q04	Medicine doses	0.88150	0.37805
PM192Q01T	Containers	0.78118	0.43469
PM992Q01	Spacers	0.32648	0.74406
PM447Q01	Tile arrangement	0.30008	0.76411
PM474Q01	Running time	0.24695	0.80494
PM564Q02	Chair lift	0.07457	0.94056
PM603Q01T	Number check	-0.07507	0.94016
PM953Q04D	Flu test	-0.09105	0.92745
PM953Q03	Flu test	-0.31919	0.74958
PM496Q02	Cash withdrawal	-0.86141	0.38901
PM446Q02	Thermometer cricket	-0.91201	0.36176
PM909Q02	Speeding fines	-1.41470	0.15716
PM034Q01T	Bricks	-1.48502	0.13754
PM00FQ01	Apartment purchase	-1.49878	0.13393
PM909Q03	Speeding fines	-1.60461	0.10858
PM423Q01	Tossing coins	-1.66780	0.09536
PM155Q03D	Population pyramids	-1.68672	0.09166
PM559Q01	Telephone rates	-1.85992	0.06290
PM982Q03T	Employment data	-1.89390	0.05824
PM00GQ01	An advertising column	-1.92636	0.05406
PM919Q01	Z's fan merchandise	-2.15367	0.03127
PM998Q04T	Bike rental	-2.24063	0.02505
PM992Q03	Spacers	-2.35879	0.01833
PM923Q04	Sailing ships	-2.37694	0.01746
PM923Q03	Sailing ships	-2.38357	0.01715
PM995Q02	Revolving door	-2.41466	0.01575
PM992Q02	Spacers	-2.46624	0.01365
PM408Q01T	Lotteries	-2.49105	0.01274
PM982Q04	Employment data	-2.55656	0.01057
PM923Q01	Sailing ships	-2.93591	0.00333
PM828Q03	Carbon dioxide	-3.18858	0.00143
PM955Q03	Migration	-3.87156	0.00011
PM949Q01T	Roof truss design	-4.42231	0.00001
PM995Q01	Revolving door	-4.81654	0.00000
PM903Q01	Drip rate	-4.85138	0.00000
PM906Q01	Crazy ants	-5.01058	0.00000
PM943Q02	Arches	-5.17043	0.00000
PM924Q02	Sauce	-5.26597	0.00000

Table 21. Overview of the Waldtest statistics for DIF-detection sorted by size (continued)

Item code	Item name	z-statistic	p-value
PM406Q01	Running tracks	-5.64003	0.00000
PM406Q02	Running tracks	-5.87354	0.00000
PM905Q02	Tennis balls	-6.60384	0.00000
PM305Q01	Map	-6.95330	0.00000
PM00KQ02	Wheelchair basketball	-7.33787	0.00000
PM464Q01T	The fence	-7.49525	0.00000
PM955Q02	Migration	-8.30818	0.00000
PM800Q01	Computer game	-8.36962	0.00000
PM915Q02	Carbon tax	-11.68626	0.00000

Note: Items that indicate a strength of the United States are highlighted in grey, whereas items that indicate a weakness are highlighted in blue.

Source: Computed using the open source data from OECD (2012a), Cognitive item response data file, www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm, www.oecd.org/pisa/pisaproducts/INT_COG12_DEC03.zip.