Comparison of Country Level Results

This chapter focuses on differences in the patterns of student performance by aspects of mathematical content contained within PISA 2003 assessment items' expectations. In participating countries, by the age of 15, students have been taught different subtopics from the broad mathematics curriculum. The subtopics vary in how they are presented to the students depending on the instructional traditions of the country.

INTRODUCTION

The PISA 2003 assessment framework (OECD, 2003) emphasises that "*mathematical literacy* focuses on the capacity of 15-year-olds (the age when many students are completing their formal compulsory mathematics learning) to use their mathematical knowledge and understanding to help make sense of [issues affecting engaged citizens in their real-worlds] and to carry out the resulting tasks" (p. 24). However, the amount and content of this knowledge 15-year-olds hold is largely dependent on what they have learnt at school. This learning appears to vary greatly across schools between and within countries.

Differences in curricula and traditions …

Evidence from the Third International Mathematics Study (TIMSS) that "the countries' traditions in mathematics education placed unequal emphasis on these subtopics in the curriculum, and as a consequence of this the students' performances were also quite different" (Zabulionis, 2001). Similar results regarding patterns of performance were obtained by Wu (2006) in relation to PISA 2000. That is, countries with similar mathematics curriculum also had similar response patterns in assessments of student capabilities within mathematics.

The substantive analysis for this chapter begins with an investigation of the patterns of performance by country. Then, the relative difficulty of particular topics and individual items is examined after an adjustment that places each country's overall mean difficulty to 0 in order that comparisons can be made among countries.

… and in grade level partly explain performance patterns across countries …

The report continues with grade level differences in performance. Due to the differences in grade level, the knowledge accumulated by age 15 can be quite different even in the same country. This also can influence students' performance. Previous research related to TIMSS indicated that there were significant differences between countries in some topics depending on whether the topic had been taught or not (Routitsky and Zammit, 2002).

Similar results can be found in PISA depending on the country and year level of students. This chapter examines a breakdown of students' performance by country, mathematics topic and grade level to investigate the impact of curriculum and instructional traditions on the patterns of performance.

… competency clusters and context areas.

The chapter concludes with examination of item difficulty by competency clusters and context areas by country and overall. While all competency clusters are important for *mathematical literacy*, it is equally important to balance instruction in terms of difficulty. There is a wide-spread belief that some contexts are more relevant for students than others, PISA provides rich data for examination of this subject.

CROSS COUNTRY DIFFERENCES IN CURRICULUM

Using the TIMSS 1996 test item data, Zabulionis (2001) categorised participating countries' achievement patterns into four groups, using a hierarchical cluster analysis. The four groups that resulted were characterised as follows:

Previous research identifies four groups of countries with similar performance.

- English-Speaking Group: Australia, Canada, England, Ireland, New Zealand, Scotland and the United States.
- Post-Communist Group: Bulgaria, Czech Republic, Hungary, Latvia and Lithuania.
  The Russian Federation, Romania, Slovak Republic and Slovenia.
- Nordic Group: Denmark, Iceland, Norway and Sweden.
- Eastern Asian Group: Hong Kong-China, Japan, Korea, and Singapore.

This chapter attempts to link country performance, at the individual item level or for groups of items, to mathematics instruction in the countries participating in PISA 2003. By examining relative differences between countries in performance related to recognisable subtopics of mathematics, one can identify potential *mathematical literacy* weaknesses and strengths within each country. This information can provide valuable feedback for curriculum design and instructional practices. It should be noted that the comparisons carried out in this chapter focus on relative differences in performance in subtopics of mathematics within each country. That is, regardless of how well a country performed overall, relative weaknesses and strengths are identified within each country. In this way, the comparisons across countries are not simply based on a horse-race ranking of countries. Rather, the comparisons use yardsticks within each country for reference. For high performing countries, there may still be room for improvement in striking a balance in curriculum design. For low performing countries, specific areas of mathematics may be identified as trouble spots. In this way, the PISA survey can provide information beyond a simple ranking of countries, and, in doing so, relate PISA 2003 findings to potential improvements in instructional practices unique to the countries.

In addition to identifying differential performance across countries, differential performance at the item level between adjacent grades within each country may also reveal defining features or deficits in their mathematics curriculum structure. Previous research related to TIMSS indicated that there were significant differences between countries in some topics depending on whether the topic had been taught or not taught (Routitsky and Zammit, 2002). Such differences, if found in the PISA survey, could provide insight into the relationship of instruction to student performance. For example, if a country's curriculum has

a spiral structure (mathematics topics are repeated from one grade to another, but with further extensions at a higher grade level), and then one might expect smaller differences between performances across grades, at the item level. But if the curriculum has a linear structure (different topics are taught at different grade levels), then one might expect greater differences in performance across different grades, at the item level, especially if little use of the content is made in the grade level in which students are assessed.

GROUPINGS OF COUNTRIES BY PATTERNS IN ITEM RESPONSES

To identify the extent to which groups of countries have similar patterns of item responses, an analysis was carried out to obtain the relative difficulties of PISA 2003 mathematics items within each country. For example, if two countries have similar curricula, one would expect the relative item difficulties to be similar for these two countries. So, if item A is more difficult than item B for Country 1, then item A is expected to be more difficult than item B for Country 2, even if the overall performance of Country 1 is a great deal higher (or lower) than that of Country 2. Such a comparison of relative item difficulty can be carried out by comparing separately calibrated item parameters (*e.g.* difficulty, discrimination, guessing) for each country, where the mean of the item difficulties for each country is set to zero. In this way, comparisons of relative item difficulties can be made between countries, without being confounded by the overall ability of the students in each country. When an item appears to be more (or less) difficult for students in some countries than for students in other countries, it is said that the item exhibits differential item functioning (DIF) with respect to the variable "country".

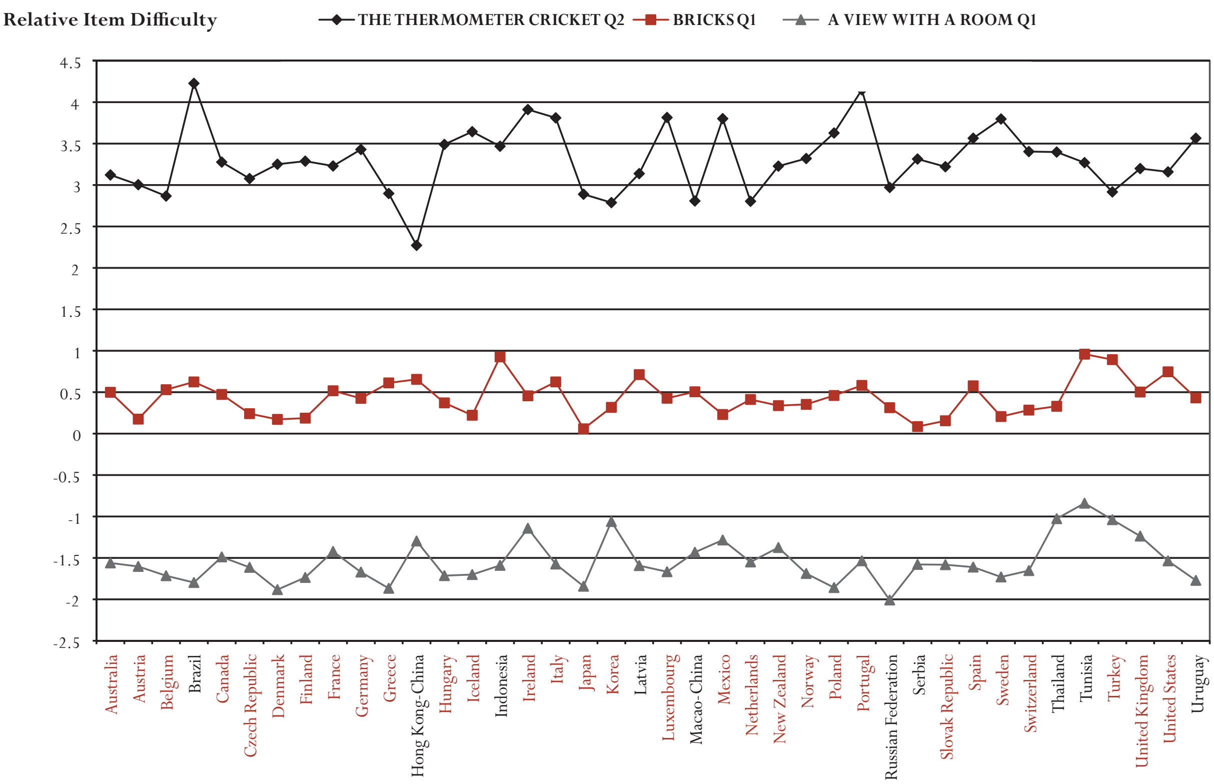
Performance patterns in PISA questions can provide useful insights for curriculum design.

In general, one would expect difficult items to be difficult for most countries, and easy items to be easy for most countries. Figure 4.1 shows a plot of item parameters by country, for three selected items. It can be seen that, overall, a difficult item (*e.g.* THE THERMOMETER CRICKET Q2) is difficult for all countries, and an easy item (*e.g.* A VIEW WITH A ROOM Q1) is easy for all countries. BRICKS Q1 was slightly above average difficulty across all of the countries. However, for each item, there are small variations among countries. These variations are expected, given that there are differences between countries in language, culture, curriculum structure, teaching methodology, and many other factors. But what is interesting is that there do appear to be patterns of groupings of countries that exhibit the same variation in the item parameters. For example, Hong Kong-China, Macao-China and Korea all found THE THERMOMETER CRICKET Q2 relatively easier as compared to other countries. In contrast, Brazil and Portugal found this item more difficult as compared to other countries. The question THE THERMOMETER CRICKET Q2 requires students to understand relationships between variables and to express the relationships algebraically.

It is interesting to study each item and the skills required for the item, and note each country's relative performance on the item, and the extent to which the item exhibits DIF. Unfortunately, such a study will not be overly useful in informing mathematics teaching in general, as the results relate only to isolated items. However, on the other hand, analyses of groups of countries with similar patterns of calibrated item parameters have the potential to provide more powerful information on which to form hypotheses about curriculum structures within groups of countries.

To identify groups of countries with similar patterns of performance, a hierarchical cluster analysis[1] was carried out on the separately calibrated item parameters for countries and sub-regions. Sub-regions with different languages are included in the cluster analysis to provide some information on the importance of language on student performance results. Figure 4.2 shows the dendrogram generated by the cluster analysis. This diagram shows from bottom to top the order in which similar countries join together in "shortest distances" between the joining countries in terms of the patterns of item difficulty parameters. For example, Australia and New Zealand are the two closest countries in terms of their patterns of item difficulties. They are then joined by the United Kingdom, and then joined by Canada (English) and Scotland.

Figure 4.1 ■ **Comparison of item parameters by countries for three selected items**



Performance in PISA suggests the following two large groups of countries: i) English speaking countries (except the United States) and Scandinavian countries

ii) European countries.

The results from the cluster analysis can be summarised as follows (and also shown by visually grouped countries in Figure 4.2). Apart from the United States, English-speaking countries form a cluster grouping with similar performance patterns. They are joined by Scandinavian countries, however the Scandinavian countries (Finland, Denmark, Norway and Sweden) are more similar to the English-speaking countries than they are to each other. Further, the Scandinavian countries are closer to English-speaking countries than to other European countries.

European countries form a cluster, and in particular, countries sharing the same language tend to have very similar performance patterns. For example, Germany and the German-speaking part of Luxembourg, Italy and the Italian-speaking part of Switzerland, Austria and the German-speaking part of Switzerland, etc., all show close links with each other in their patterns of item difficulties. Interestingly, Eastern European countries such as the Czech Republic, the Slovak Republic, Hungary, Poland and the Latvian-speaking part of Latvia all show closer links to Western European countries than to the Russian Federation, a result that is different from the findings of earlier studies such as the one carried out by Zabulionis (2001).

East Asian countries such as Japan, Hong Kong-China and Korea are somewhat different from each other, as well as different from English or European groups. Note, however, that Hong Kong-China and Macao-China are very closely linked, and Japan and Korea are closer together than they are to other countries.

It is difficult to clearly identify the factors accounting for the observed clusters, since language, culture, geographical locations and educational traditions are so intertwined that it is nearly impossible to clearly separate the four. Some may suggest that the clusters are simply formed by language groups. This is not quite right. For example, the French-speaking part of Canada is closer to the English-speaking countries and Brazil is closer to Mexico than to Portugal.

While the underlying reasons for the observed clustering may be difficult to identify, the results of the cluster analysis provide us with a starting point for making further hypotheses and investigation.

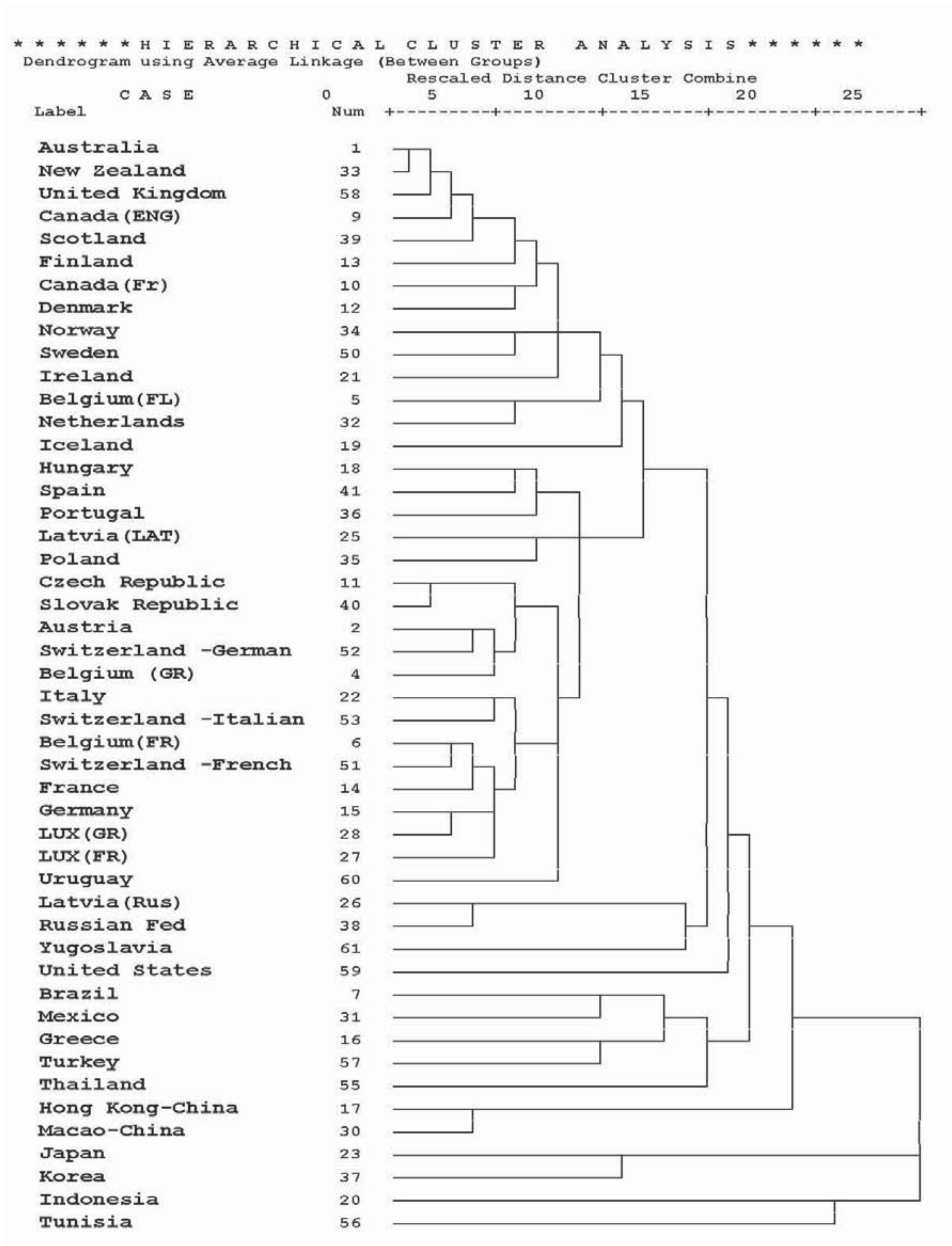
PATTERNS IN MATHEMATICS CONTENT

PISA questions can be classified by traditional mathematic domains.

Given that some countries found particular items more (or less) difficult than other countries, it would be interesting to examine whether there exist patterns in the DIF results for each mathematics topic. For example, a country may perform consistently better (or worse) on a particular mathematics topic relative to other topics. Or, perhaps, groups of countries may show the same pattern across different mathematics topics, depending on the way mathematics is taught in the countries.

For such an analysis to be carried out, PISA mathematics items first need to be classified according to traditional curriculum topics. Analyses can then be carried out for specific mathematics topics and for specific groups of countries. In this study,

Figure 4.2 ■ **Hierarchical cluster analysis of item parameters**



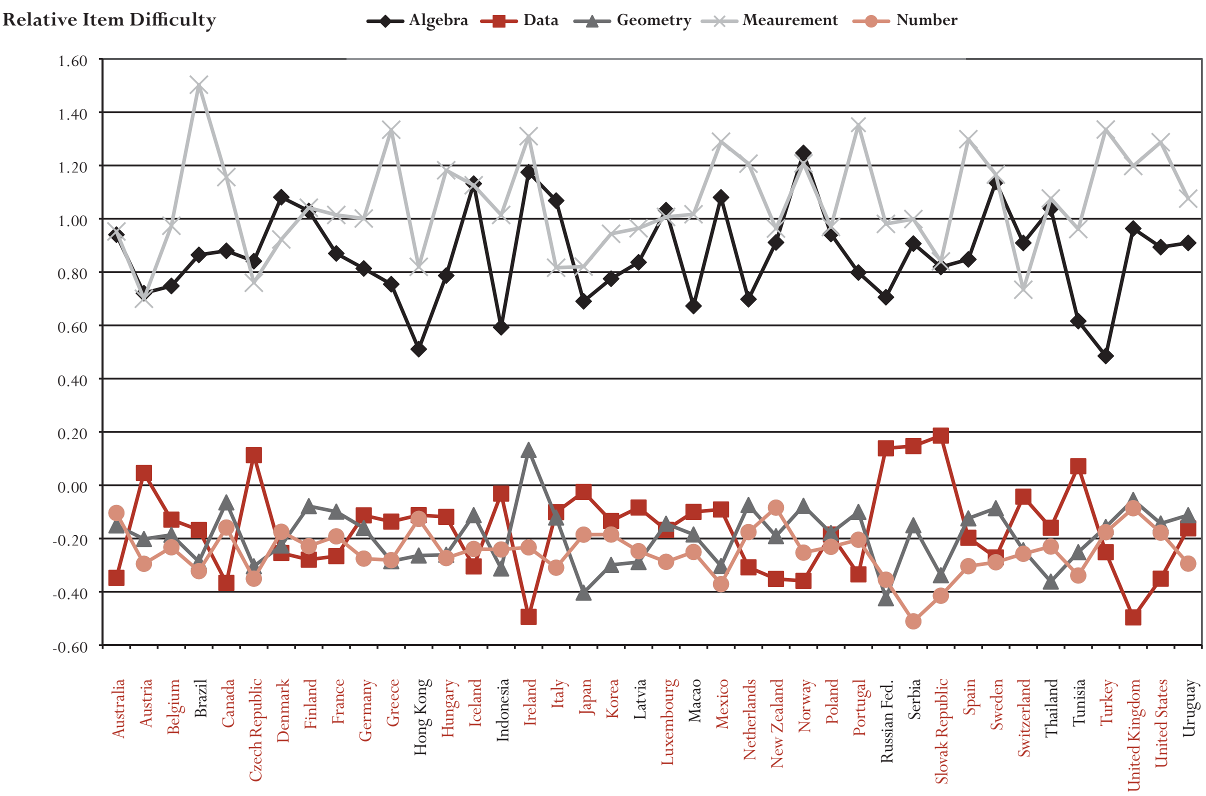
the PISA mathematics items have been classified under the following five general mathematics curriculum topics: Number, Algebra, Measurement, Geometry and Data. These five curriculum topics are typically included in national curriculum documents in many countries. They also match the TIMSS framework classifications. Consequently, the information collected in the Test-Curriculum Match Analysis (TCMA) in TIMSS (see Chapter 5 of *TIMSS 2003: International Mathematics Report* [Mullis, Martin, Gonzalez, and Chrostowski, 2004]) can also be used as supporting evidence to link curriculum to test results.

For each country, item difficulties for items classified under the same mathematics topic are averaged to provide an indication of the level of difficulty of each mathematics topic in each participating country, relative to the difficulty of other mathematics topics in the same country.

Figure 4.3 shows the average item difficulty for each mathematics topic in each country. It is important to note that the overall mean difficulty for each country has been set to 0 to allow for comparisons among nations.

The data pictured in Figure 4.3 indicate that Algebra and Measurement items are generally more difficult than Data, Geometry, and Number items in all

Figure 4.3 ■ **Relative difficulty by mathematics topic by country**[1]



1. Please note that mean difficulty for each country across all mathematics items is set to 0.

countries. Moreover, there is little variation across countries in average item difficulty for Number items. Greater variability is observed across countries in average item difficulty for Algebra, Measurement, and Data items. This is expected, as every country covers Number topics in earlier grades of schooling, while there are greater differences in the grade levels at which Algebra, Measurement, and Data are introduced and taught. The mean and the standard deviation of topic item difficulty across countries are given in Table 4.1.

To make grouping of the countries clearer, the relative easiness/difficulty of the topics within the countries are calculated. The topic is defined being relatively difficult (D) if the average difficulty for the country illustrated in Figure 4.3 is half a standard deviation (or more) larger than the mean across the countries provided in Table 4.1. Similarly, the topic is defined being relatively easy (E) if the average difficulty for the country illustrated in Figure 4.3 is half a standard deviation (or more) smaller than the mean across the countries provided in Table 4.1.

Algebra and Measurement are relatively more difficult, while Data, Geometry and Number are easier.

Table 4.1
Mean and standard deviation of relative topic difficulty across countries

Topic	Topic difficulty across countries Mean (SD) in logits
Number (32 items)	-0.25 (0.09)
Algebra (7 items)	0.87 (0.18)
Measurement (8 items)	1.06 (0.19)
Geometry (12 items)	-0.18 (0.12)
Data (26 items)	-0.17 (0.17)

Table 4.2 provides a summary of the four groups of countries identified at the beginning of this chapter. The full information about relative difficulty of the traditional topics can be found in Annex A3.

Across domains, one can identify different country groups.

Table 4.2 shows that for all English-speaking countries the topic Data is a relatively easy topic and for all of them, except Ireland, the topic Number is a relatively difficult topic. As in the factor analysis, Australia and New Zealand are more similar to each other than to other English-speaking countries. The United Kingdom is similar to Ireland on the one hand and to Canada on the other.

For example, data is relatively easy for English-speaking countries …

For the Northern European countries Table 4.2 also provides further insight in where exactly the similarities lie. For all of them Algebra is a relatively more difficult topic. For all except Denmark, Data is relatively easy and Geometry is relatively difficult. Noticeably, the only similarity Denmark shows with other Scandinavian countries at the topic level is the relative difficulty of Algebra.

Table 4.2
Relative easiness/difficulty of each topic within the countries

Country	Algebra	Data	Geometry	Measurement	Number
Australia		E		E	D
New Zealand		E		E	D
Canada		E	D	D	D
United Kingdom	D	E	D	D	D
Ireland	D	E	D	D	
United States		E		D	D
Finland	D	E	D		
Norway	D	E	D	D	
Sweden	D	E	D	D	
Iceland	D	E	D		
Denmark	D			E	D
Japan	E	D	E	E	D
Korea	E		E	E	D
Hong Kong-China	E		E	E	D
Macao-China	E				
Yugoslavia		D			E
Russian Federation	E	D	E		E
Latvia		D	E	E	
Czech Republic		D	E	E	E
Slovak Republic		D	E	E	E
Austria	E	D		E	E
Switzerland		D		E	

For the South-East Asia region, Algebra is a relatively easy topic and Number (except Macao-China) is a relatively difficult topic. Other traditional topics like Geometry and Measurement are relatively easy for three out of four South-East Asian countries, with Macao-China being the exception.

... but data is relatively difficult for Central European countries.

Some of the Central European countries and all post-communist countries share the relative difficulty of the topic Data. Further, with the exception of the Russian Federation and Yugoslavia, the topic of Measurement is comparatively easy. For the Russian Federation, Latvia, the Czech Republic and the Slovak Republic the topic of Geometry is relatively easy. Clearly, the Czech Republic and the Slovak Republic are more similar to each other than to other countries in the group and this finding was confirmed by the factor analysis discussed at the beginning of this section.

Going deeper into the curriculum structure for Data provides more insights ...

The association between curriculum coverage and performance on mathematics topics is further explored by examining curriculum structure at the country level. The mathematics topic Data has been chosen for a detailed discussion below, mainly because there are variations across countries in terms of the grades at which Data topics are taught, as well as the availability of curriculum information for some countries. Figure 4.3 shows that the eight countries where students found the Data topic relatively more difficult are the Slovak Republic, Serbia, the Russian Federation, the Czech Republic, Tunisia, Austria, Japan and Indonesia. In contrast, the eight countries where students found the Data topic relatively easier are the United Kingdom, Ireland, Scotland, Canada, Norway, Australia, New Zealand and the United States. For some of the countries the grades at which Data is taught can be found from the data collected in TIMSS TCMA (Mullis, Martin, Gonzalez, and Chrostowski, 2004). Table 4.3 shows the relative performance of these countries in Data (in relation to their performance on other mathematics topics) and corresponding curriculum information.

Unfortunately, data are not available as part of the TCMA for Ireland, Canada, the Czech Republic and Austria. With the six remaining countries having relatively high and relatively low difficulty indices for data items, interesting patterns emerge. The data in Table 4.3 are entered by grade levels where the content associated with data are usually focused on within a country's curriculum. In some cases, especially where there is national guidance, the grade levels are entered. In other cases, where the control on focus and degree of emphasis is handled at regional or local levels, the emphasis is indicated by whether all or most students, indicated by a Y, have received this coverage by the end of Grade 8, the most able students have received it by the end of Grade 8, indicated by an M, or this has not been included for study by the end of Grade 8, indicated by an N.

The columns in the table, labelled by Roman numerals, are associated with specific Data related outcomes:

I. Organising data using one or more characteristics using tally charts, tables or graphs

II. Sources of errors in collecting and organising data

III. Data collection methods

IV. Drawing and interpreting graphs, tables, pictographs, bar graphs, pie charts and line graphs

V. Characteristics of data sets including mean, median, range and shape of distribution

VI. Interpreting data sets

VII. Evaluating and interpreting of data with respect to correctness and completeness of interpretation

VIII. Simple probabilities including using data from experiments to estimate probabilities for favourable events

Table 4.3
Average item difficulty parameter values for Data items[1]

		Data topics (TIMSS TCMA topics)							
		I	II	III	IV	V	VI	VII	VIII
	Country	Grades at which the topic is taught							
Less difficult	United Kingdom[2]	K-6	8-10	6-10	6-10	5-10	6-10	8-10	4-8
	Scotland	Y[3]	M[4]	Y	Y	M	Y	M	M
	Norway	7-10	8-10	8-10	6-10	8-10	8-10	8-10	9-10
	Australia	4-7	9-10	5-8	5-8	8	9-10	9-10	7-10
	New Zealand	3-9	6-9	6-9	3-9	6-9	6-10	8-9	5-9
	United States	Y	Y	Y	Y	Y	Y	Y	Y
More difficult	Indonesia	8-9	10	11	8-11	10-12	11	11	10
	Japan	3-5	10-12	10-12	3-5	10-12	10-12	10-12	8
	Tunisia	N[5]	N	N	N	N	N	N	N
	Russian Federation	N	Y	N	Y	N	N	N	N
	Serbia	8	12	12	12	12	12	12	12
	Slovak Republic	9	9	9	7	8	9	9	7

1. From TIMSS TCMA results.
2. Note that in TIMSS, the TCMA only had data on England. In PISA, the data collected were for the United Kingdom.
3. Note that "Y"indicates that nearly all students received coverage of this topic by the end of Grade 8.
4. Note that "M"indicates that the most able students received coverage of this topic by the end of Grade 8.
5. Note that "N"indicates that the topic was not included by students by the end of Grade 8.

The first observation to be made about Table 4.3 is that the countries that find Data easier are predominantly Western countries, seven of which are English-speaking. These countries could be regarded as having similar educational traditions. In contrast, the countries that find Data more difficult are non-Western countries, three of which are from Eastern Europe where these countries also have similar educational traditions.

… for example: countries where Data is relatively easier, introduce the topic from an earlier grade.

The second observation to be made about Table 4.3 is that the countries finding Data easier tend to teach the topic from an earlier grade. For example, in England, all of the eight Data topics are introduced at the K-8 level in schools. In New Zealand, seven Data topics are introduced in primary schools. In contrast, in Serbia, only one data topic is introduced by the intermediate level and none of the topics are introduced at this level in Tunisia.

The third observation about Table 4.3 is that the countries where the Data topic is relatively more difficult appear to adopt a more linear organisational structure of the mathematics curriculum, where specific topics are only taught at specific grade levels. While there is some evidence of the same case in the countries where the Data topics were relatively less difficult, the length of the intervals of focus appear to be slightly longer. This may suggest a spiral organisational structure of curriculum, where each curriculum topic is taught across many year levels (*e.g.* 3 to 5 year levels). Consequently, in Japan, for example, students who were taught Data Topic I in Grades 3 to 5 may have forgotten about this content domain by the time they reach age 15 (Mullis, Martin, Gonzalez, and Chrostowski, 2004).

The analyses carried out in this section provide some evidence linking student performance to instruction. Countries where students have received more instruction on a mathematics topic tend to perform better in that topic relative to their performance on other topics.

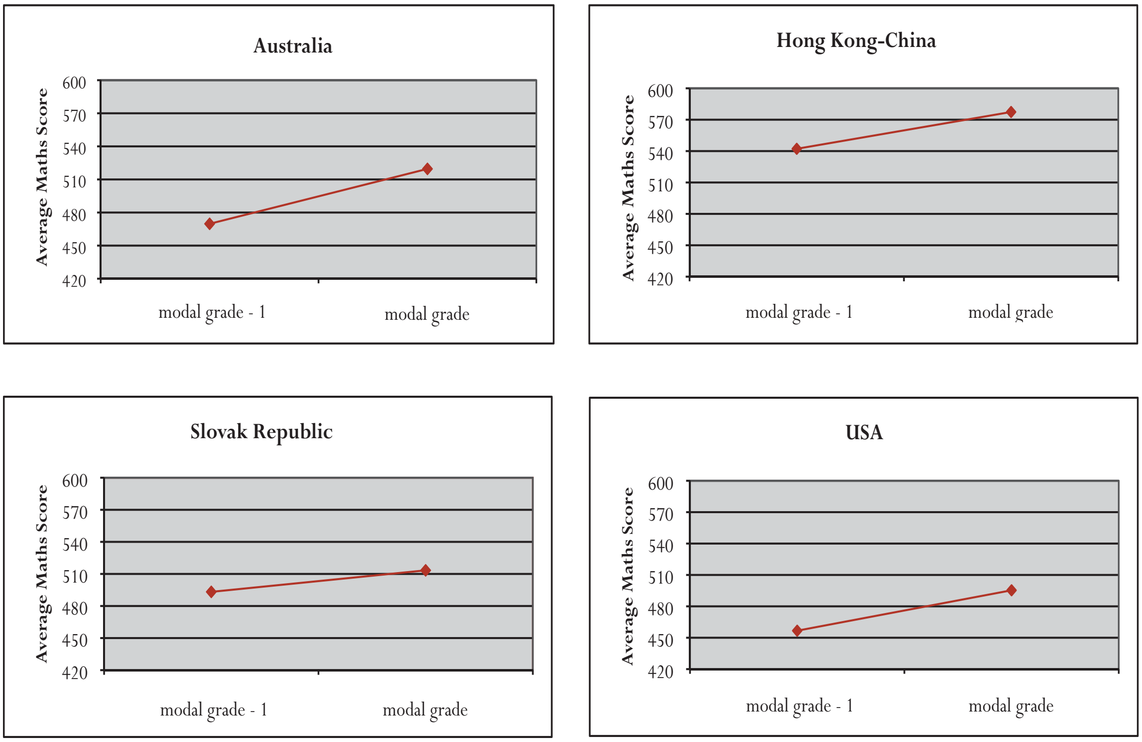
PERFORMANCE AND GRADE LEVELS

Students from higher grades perform better.

If instruction has a significant impact on student performance, then one would expect some differences in performance between students from different grade levels. Clearly, one would expect students from higher grade levels to perform better, on average, than students from lower grade levels. Figure 4.4 shows the relative performance of students from different grades for a number of countries randomly selected for the purpose of illustration.

Generally speaking, for all countries that have multiple grades in the PISA study, students from higher grades performed, on average, better than students from lower grades although the magnitudes of the differences in performance between adjacent grades varied between countries. This, of course, is expected, as one additional year of schooling must increase students' performance level. Nevertheless, this finding offers further evidence that instruction is closely related to performance.

Figure 4.4 ■ **Average performance by grade for four participants**



An analysis of the performance differences across grades shows that student performance is closely related with instruction.

Whether a country adopts a spiral structure of curriculum or a linear structure, there will be some variations in the topics taught at each grade level. Consequently, some items may show Differential Item Functioning (DIF) for students because of the inclusion of students from different grade levels. That is, for some items, students with the same ability will be likely to have different probabilities of success if they are from different grade levels. To test this hypothesis, DIF analyses were carried out for a selected number of countries where there were substantial numbers of students from different grades. The items for which lower grade students were most disadvantaged were identified. These items were further examined in terms of content and, where possible, in terms of national curriculum. Table 4.4 shows the results.

A number of observations can be made from the analyses of Grade DIF. First, the magnitudes of DIF across grades within countries are generally less than the magnitudes of DIF across countries. For a particular item, the maximum difference between item difficulty parameters across countries is typically between 1 and 2 logits (see variations of item difficulties in Figure 4.1, for example). In comparison, within a country, the maximum difference between item difficulty parameters between two grades is around 0.5 logit.

Second, of the eighteen entries in Table 4.4, eight are Algebra, although only four of these eight entries are different items. In particular, the Algebra item THE

THERMOMETER CRICKET Q2 [M446Q02] appears in four of the six countries, as showing differential item functioning between two adjacent grades. This item is also the most difficult item among all mathematics items in PISA 2003. As Algebra is generally not taught until later years of schooling, it is not surprising that there is a greater chance that these items show DIF across grade levels. That is, when there are two students with the same overall mathematics ability, the student from a higher grade will have a higher probability of success on an algebra item than a student from a lower grade. From this point of view, the

Table 4.4
Items identified with grade DIF for countries with multiple grades

	Three items where lower grade is disadvantaged most	Mathematics topic	Subtopic	Grades taught	Item difficulty calibrated for the country (logits)	Item difficulty difference between two grades (logits)
Australia	THE THERMOMETER CRICKET Q2	Algebra	Equations	9-10	3.22	0.46
	WALKING Q01	Algebra	Equations	9-10	1.26	0.33
	EXCHANGE RATE Q02	Number	Ratios	9-10	-1.11	0.32
Austria	RUNNING TRACKS Q03	Measurement	Formulas	N[1]	1.63	0.32
	THE THERMOMETER CRICKET Q2	Algebra	Equations	N	3.06	0.29
	WALKING Q01	Algebra	Equations	N	0.57	0.27
Hong Kong-China	GROWING UP Q3	Data	Interpretation	10-11	1.15	0.74
	THE BEST CAR Q01	Algebra	Equations	7-9	-1.92	0.34
	THE BEST CAR Q02	Algebra	Equations	7-9	1.17	0.33
Russian Federation	HEIGHT Q01	Data	Interpretation	N	-0.06	0.36
	THE THERMOMETER CRICKET Q2	Algebra	Equations	N	3.1	0.36
	CUBES Q01	Data	Represent	N	-0.55	0.31
Slovak Republic	CHOICES Q01	Number	Patterns	9	-0.12	0.32
	HEIGHT Q02	Data	Statistics	8	2.13	0.28
	THE FENCE Q01	Measurement	Formulas	9	1.34	0.28
United States	THE THERMOMETER CRICKET Q2	Algebra	Equations	N	3.26	0.5
	CUBES Q01	Data	Represent	N	-0.99	0.36
	CARBON DIOXIDE Q01	Data	Represent	N	0.62	0.36

1. Note that “N” indicates that no information is available.

students from lower grades are disadvantaged, owing to the fact that they have not received as many instructional lessons on Algebra as students from higher grades. This is often referred to as the OTL, or opportunity to learn, factor.

Third, in addition to Algebra items, some Data items also exhibit DIF across grade levels. In particular, Hong Kong-China, the Russian Federation, the Slovak Republic and the United States all have Data items showing grade DIF. In Table 4.3 students in the Russian Federation and the Slovak Republic found Data items more difficult than students in most other countries. This is an indication that Data is generally not taught until higher grades in these countries. Consequently, it is not surprising that some Data items also show grade DIF.

Fourth, for countries where there is information about curriculum structure, it appears that grade DIF items relate to topics that are taught at higher grade levels. For example, a Number item is identified as exhibiting grade DIF for Australia. This Number topic is only taught at Grade 9. Similarly, for Hong Kong-China, the Data item GROWING UP Q3 is found to exhibit grade DIF. The content area for this item is taught at Grades 10-11.

Finally, grade DIF items tend to be more difficult items. The average item difficulty for the items in Table 4.3 is 0.99 logit, where the average item difficulty for the whole set of mathematics items is 0 logit for each country.

In summary, the identification of grade DIF items provides support for the hypothesis that student performance is closely linked to instruction. Moreover, the identification of specific grade DIF items for each country can shed some light on the curriculum structure in the country and provide the basis for possible intervention strategies if necessary. However it must be noted that PISA is designed primarily as an age-based survey, so the presence of multiple grades within a country is not controlled. As such, the study design of PISA does not lend itself to in-depth analysis of grade differences for all countries.

COMPETENCY CLUSTERS AND MATHEMATICS PERFORMANCE

In PISA mathematical competencies are organised into three clusters (see Chapter 2). The PISA 2003 mathematics assessment included 26 questions in the *reproduction* competency cluster, 40 questions in the *connections* competency cluster, and 19 questions in the *reflection* competency cluster.

Across competency clusters, Reproduction was the easiest and Reflection the most difficult.

The relative difficulties of questions included within each competency cluster are presented by country in Figure 4.5. These statistics show that questions in the *reproduction* competency cluster were on average the easiest and those in the *reflection* competency cluster were on average the most difficult. This relationship for difficulty of questions within competency clusters holds for all of the participating countries.

Table 4.5
Mean and standard deviation of question difficulty by competency cluster across countries

Competency cluster	Number of questions included	Difficulty of questions included across countries (in logits) Mean (SD)
Reproduction	26	-1.00 (0.06)
Connections	40	0.26 (0.03)
Reflection	19	0.82 (0.09)

Moreover, there is little variation across countries in average question difficulty for all competency clusters. The mean and the standard deviation of question difficulty for each competency cluster across countries are given in Table 4.5.

It was shown earlier in this chapter (see Table 4.3) that Algebra and Measurement questions are significantly more difficult than Number, Geometry and Data across all countries. Table 4.6 shows the distribution of PISA questions by the traditional mathematics topic and by competency cluster. The competency clusters include questions from each of the traditional mathematics topics, although the *reproduction* competency cluster does not include Measurement questions.

Figure 4.5 ■ **Average question difficulty by competency cluster in participating countries**

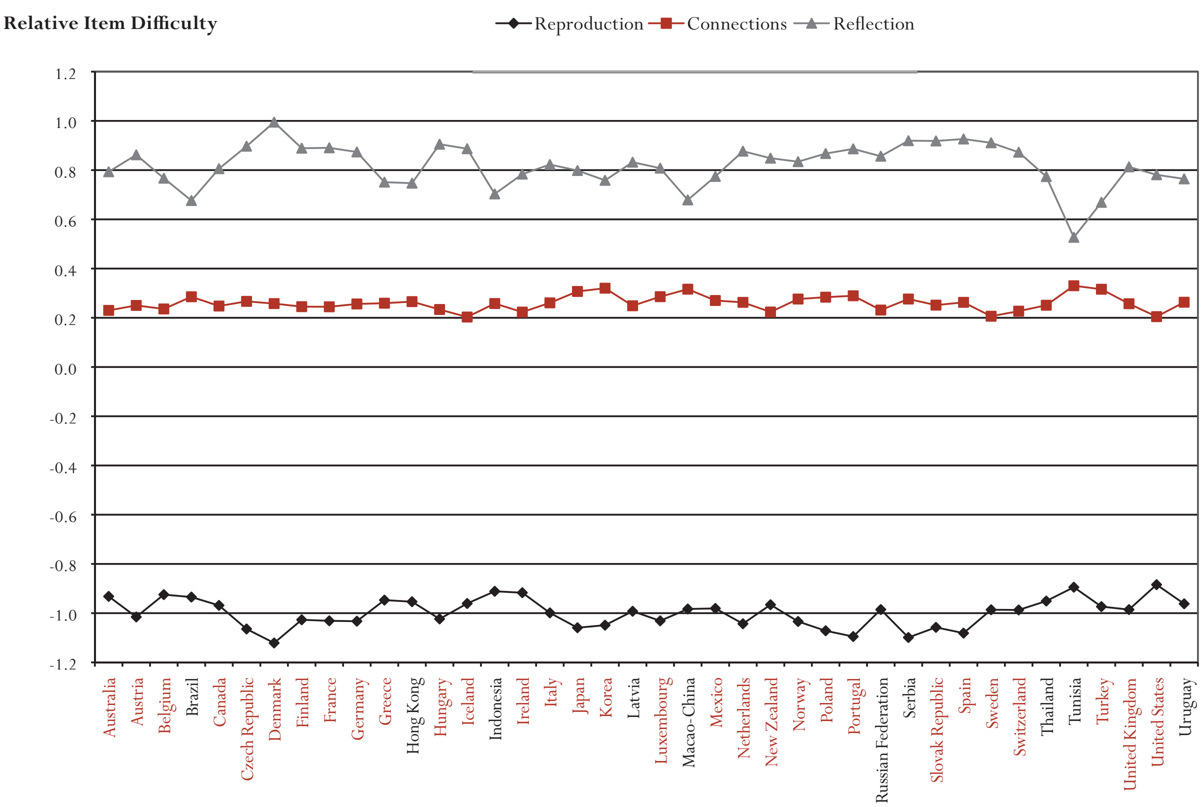


Table 4.6
Questions in competency clusters by traditional mathematics topic

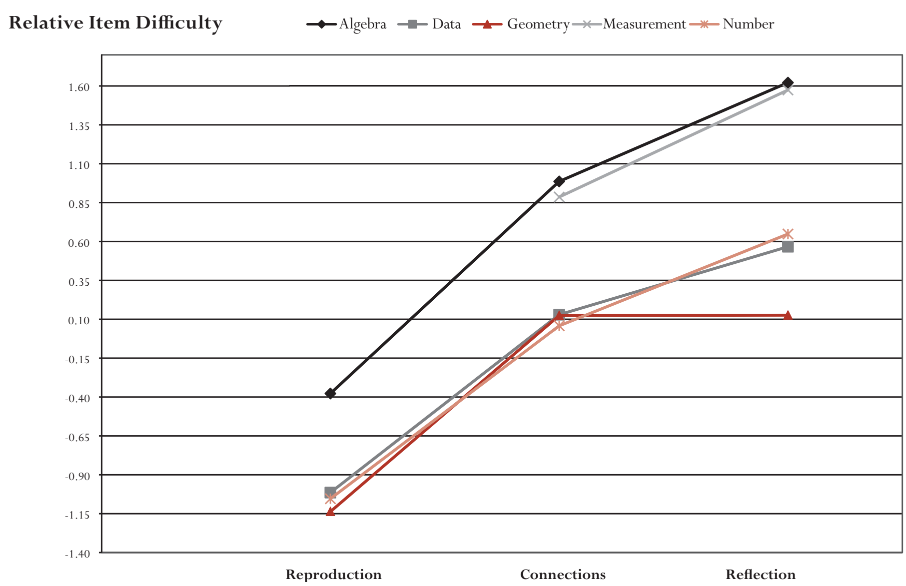
	Algebra	Data	Geometry	Measurement	Number
	Percentage of questions (number of questions)				
Reproduction	28.6% (2)	34.6% (9)	25.0% (3)	0.0% (0)	37.5% (12)
Connections	28.6% (2)	42.3% (11)	58.3% (7)	75.0% (6)	43.8% (14)
Reflection	42.9% (3)	23.1% (6)	16.7% (2)	25.0% (2)	18.7% (6)
Total	100% (7)	100% (26)	100% (12)	100% (8)	100% (32)

Both competency clusters and traditional domains are related to the difficulty of PISA questions.

Figure 4.6 suggests that the difficulty of the content in PISA questions is attributed to the traditional mathematics topic, as well as to the competency clusters. That is Algebra questions are more difficult on average within each competency cluster. The same applies to Measurement. On the other hand, the *reflection* competency cluster is more difficult within each traditional topic except Geometry. Interaction between competency clusters and traditional topics is most likely due to this Geometry effect and the absence of Measurement questions in the *reproduction* cluster.

Algebra questions are more likely to involve such competencies as symbols and formalism (see Chapter 2) which is a defining competency of *mathematical literacy*. It relates to the ability to handle and work with statements containing symbols and formulas, as for example, in THE BEST CAR – Question 1. This is the easier of two Algebra questions in the *reproduction* competency cluster and

Figure 4.6 ■ **Average question difficulty by competency cluster and by traditional topic**



requires number substitution into a given formula. The second of these questions is WALKING – Question 1 which is much more difficult as it requires both substitution into a given equation and solving the equation (see more about Algebra questions in Chapter 5).

CONTEXT AND MATHEMATICS PERFORMANCE

PISA questions presented in a personal context were easier than those in other contexts.

PISA's focus on *mathematical literacy* reflects an increasing concern about how well students can apply mathematics to solve real-life problems. Therefore, PISA mathematics questions are contextualised, reflecting different aspects of the real world such as travel, sport, media, modern communication and science, but also intra-mathematical contexts that reflect part of students' experience of mathematics in school.

PISA questions are classified into four different contexts or situations: *educational and occupational*, *scientific*, *personal* and *public* (see Chapter 2). Figure 4.7 presents the average question difficulty in each of the four contexts for each country. As in the previous sections, the mean difficulty for all mathematics items is set to 0 for all countries.

Overall, the easiest for all countries were questions presented in a *personal* context, and for the majority of countries the most difficult were questions presented in a *scientific* context. The standard deviations (see Table 4.7) are

Figure 4.7 ■ **Average question difficulty by context in participating countries**

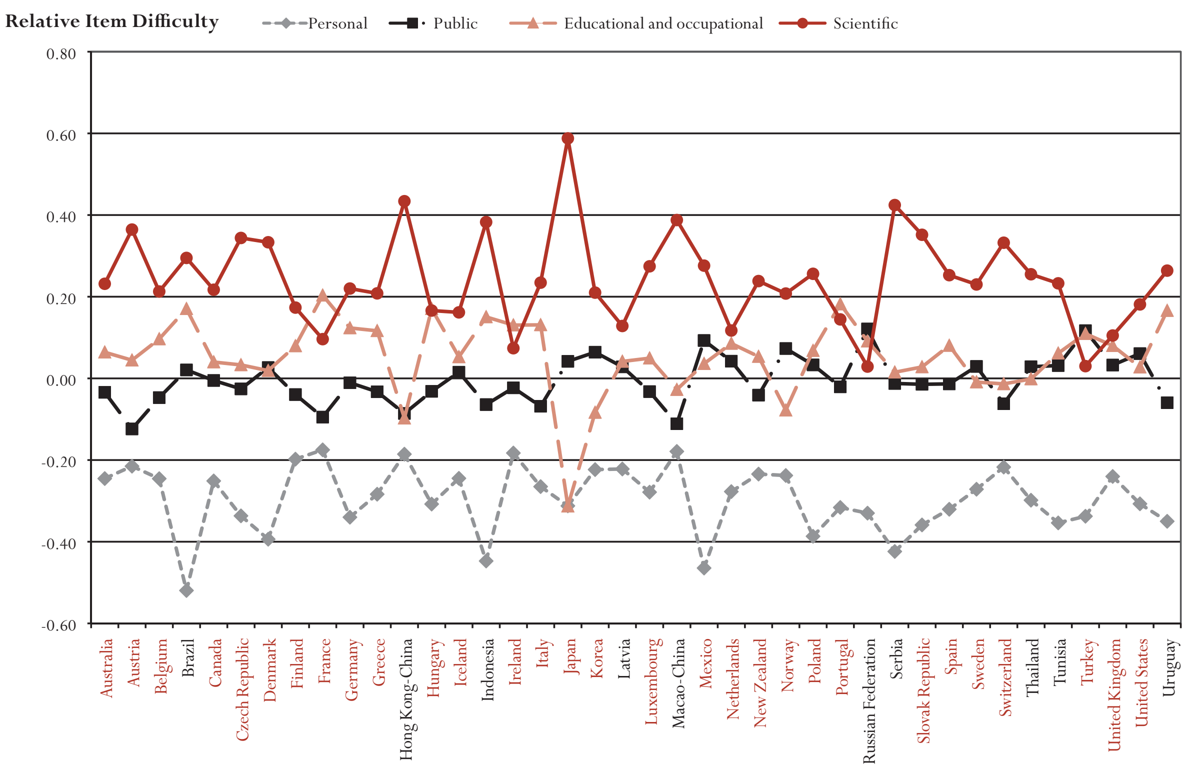


Table 4.7
Mean and standard deviation of question difficulty by context across countries

Context	Number of questions included	Difficulty of questions included across countries (in logits) Mean (SD)
Scientific	18	0.24 (0.11)
Educational and occupational	20	0.05 (0.09)
Public	29	0.00 (0.06)
Personal	18	-0.29 (0.08)

slightly higher than those observed for competency clusters but still relatively small.

Multiple comparisons confirmed that overall across 40 countries questions which were presented in a *personal* context were easier than all other questions. Also, questions presented in a *scientific* context were more difficult than all other questions except those presented in an *educational and occupational* context. The differences are small, but statistically significant. There is no difference in difficulty between questions presented in a *public* context and questions presented in an *educational and occupational* context (see Table 4.8).

Table 4.8
Multiple comparisons of question difficulty by context across countries (using Bonferroni adjustment)

(I) Context	(J) Context	Mean Difference (I-J)	Std. Error	Sig.	99% Confidence Interval	
					Lower Bound	Upper Bound
Personal	Public	-0.29	0.06	0.00	-0.48	-0.10
	Educational and occupational	-0.35	0.06	0.00	-0.55	-0.15
	Scientific	-0.54	0.07	0.00	-0.74	-0.33
Public	Educational and occupational	-0.06	0.06	1.00	-0.24	0.12
	Scientific	-0.25	0.06	0.00	-0.43	-0.06
Educational and occupational	Scientific	-0.19	0.06	0.02	-0.39	0.02

Differences in question difficulty by context are not significantly different for most countries, although, for low achieving countries questions presented in a *personal* context were relatively easier (Mexico and the partner countries Brazil and Indonesia) and for Japan questions presented in a *scientific* context were relatively more difficult.

CONCLUSION

Performance in PISA is related to curriculum and instruction.

This chapter examined the performance of countries in terms of their relative strengths and weaknesses in different traditional curriculum topics. It was found that the observed differences across countries in their performance patterns could be linked to curriculum and instruction. In particular, English-speaking countries have similar performance patterns. Students in English-speaking countries tend to perform relatively better on Data questions. Where available, evidence from TIMSS shows that instruction about Data is introduced in the early grades of schooling in these countries. In contrast, the Czech Republic, Japan, the Slovak Republic and the partner countries/economies Hong Kong-China, Serbia and the Russian Federation often appear in the same groups as better performing countries in Algebra, Geometry, Measurement and Number.

Patterns of performance emerge across countries …

While these findings are similar to those of some earlier studies (*e.g.* Zabulionis, 2001), there are some differences. In particular, Eastern European countries such as the Czech Republic, Hungary, Poland, the Slovak Republic, and the Latvian-speaking part of the partner country Latvia seem to be moving closer in their performance patterns to those of Western European countries than they were several years ago. This could reflect a gradual change in direction in curriculum structure in these countries.

… grades …

An examination of the performance patterns across grades within a country shows that there are some differences across grades, particularly for topics taught only in higher grades. However, the differences in performance patterns across grades are small as compared to performance pattern differences across countries. Nevertheless, the link between instruction and student performance is again evident.

… competency clusters …

As expected, questions from the *reproduction* competency cluster on average are easier than the questions from the *connections* competency cluster, while questions in the *reflection* competency cluster are the most difficult of the three. This is true for all countries with little variation.

… and contexts.

Regarding the context, *personal* questions are on average the easiest and *scientific* questions are on average the most difficult, and this is true for all countries, although the differences are small. The challenge for educational practitioners is to make *scientific* questions more attractive for students.

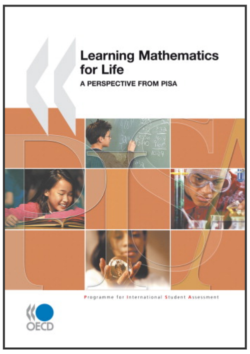
In conclusion, the analyses carried out in this chapter show that PISA results can provide useful information about student performance and instruction.

Further, curriculum structures have a significant impact on student performance. The results from this chapter provide a starting point for an examination of curriculum structures in each country, as well as an assessment of the relative merits of different curriculum designs. For example, if it is deemed important that future citizens should have a sound knowledge of statistical methods for the dissemination of information and data, then the Data topic should be introduced earlier and emphasised more in the curriculum. On the other hand, if there is a need for better preparation for tertiary science, economics, statistical, and technical studies and better understanding of future citizens of various dynamic of processes, then Algebra and the study of functions should be emphasised more. Consequently, the results from this chapter can provide a basis for a re-evaluation of curriculum designs in each country.

As PISA carries out data collection every three years, the analyses carried out in this chapter can be repeated, so that trends in the performance patterns of countries by curriculum topics can be monitored and cross-checked with curriculum changes in each country.

Note

1. For a general description of cluster analysis see, for example, Anderberg (1973).



From:

Learning Mathematics for Life

A Perspective from PISA

Access the complete publication at:
https://doi.org/10.1787/9789264075009-en

Please cite this chapter as:

OECD (2010), “Comparison of Country Level Results”, in *Learning Mathematics for Life: A Perspective from PISA*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264075009-6-en