Mathematical Problem Solving and Differences in Students' Understanding

This chapter concentrates on problem solving methods and differences in students' mathematical thinking. It discusses the processes involved in what is referred to as the "mathematisation" cycle. The chapter provides two case studies, explaining how the elements required in the different stages of mathematisation are implemented in PISA items.

INTRODUCTION

In problem-solving students apply their mathematical literacy using different methods and approaches.

PISA 2003 made a special effort to assess students' problem solving, as this is where *mathematical literacy* has its real application in life. The correlation between students' performance on overall mathematics items and their performance on those specifically focusing on problem solving was 0.89, which is higher than the correlation between mathematics and science (0.83). Nevertheless, analyses of assessment results on problem solving showed that students doing well in problem solving are not simply demonstrating strong mathematical competencies. In fact, in many countries students perform differently in these two domains (OECD, 2004b).

This chapter explains how mathematical problem-solving features are revealed in PISA questions. The PISA 2003 assessment framework (OECD, 2003) gives rise to further possibilities for investigating fundamentally important mathematical problem-solving methods and approaches. In particular, the framework discusses processes involved using the term mathematisation. The scoring design of PISA 2003 mathematics questions does not always allow for a full study of the patterns in students' responses in relation to their mathematical thinking; nevertheless, the discussion of the questions where the full problem-solving cycle comes alive can be useful for instructional practices.

PISA can also be used to analyse student strategies and misconceptions.

One area of the analysis of PISA items of particular interest to mathematics educators is the focus on student strategies and misconceptions. Misconceptions, or the study of students' patterns of faulty performances due to inadequate understandings of a concept or procedure, are well documented in the mathematics education literature (Schoenfeld, 1992; Karsenty, Arcavi and Hadas, 2007). Although PISA was not set up to measure misconceptions, the use of double scoring of some of the PISA items and the particular focus of others allow for findings of instructional interest to mathematics educators.

GENERAL FEATURES OF MATHEMATICAL PROBLEM SOLVING IN PISA

Mathematisation refers to the problem-solving process students use to answer questions.

The section begins with description of the "problem-solving process" or the process of "mathematisation" as it is called in the PISA framework of mathematical literacy (OECD, 2003). Two case studies of PISA questions that make the problem-solving cycle visible are then presented.

The mathematisation cycle …

The "problem-solving process" is generally described as a circular process with the following five main features:

1. Starting with a problem based in a real-world setting.
2. Organising it according to mathematical concepts and identifying the relevant mathematics.
3. Gradually trimming away the reality through processes such as making assumptions, generalising and formalising, which promote the mathematical

features of the situation and transform the real-world problem into a mathematical problem that faithfully represents the situation.

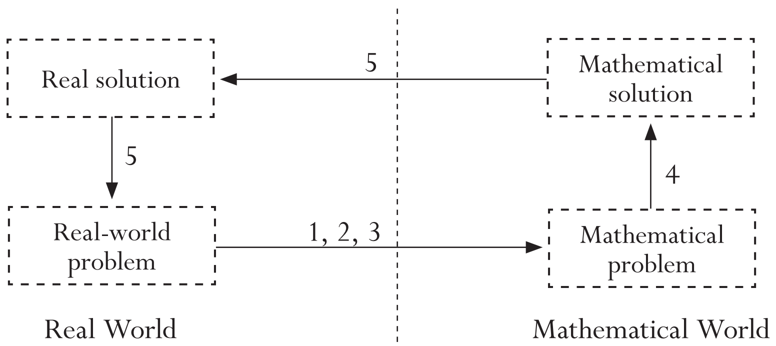
4. Solving the mathematical problem.

5. Making sense of the mathematical solution in terms of the real situation, including identifying the limitations of the solution.

Figure 6.1 shows the cyclic character of the mathematisation process.

The process of mathematisation starts with a problem situated in reality (1).

Figure 6.1 ■ **Mathematisation cycle**



Next, the problem-solver tries to identify the relevant mathematics and reorganises the problem according to the mathematical concepts identified (2), followed by gradually trimming away the reality (3). These three steps lead the problem-solver from a real-world problem to a mathematical problem.

The fourth step may not come as a surprise: solving the mathematical problem (4).

Now the question arises: what is the meaning of this strictly mathematical solution in terms of the real world? (5)

... and the three phases of mathematisation.

These five aspects can be clustered into three phases according to general features of mathematical problem-solving approaches (see, for example, Polya, 1962; and Burkhardt, 1981):

Phase 1. Understanding the question (*e.g.* dealing with extraneous data), which is also called horizontal mathematisation.

Phase 2. Sophistication of problem-solving approaches, which is also referred to as vertical mathematisation.

Phase 3. Interpretation of mathematical results (linking mathematical answers to the context).

MAKING THE PROBLEM-SOLVING CYCLE VISIBLE THROUGH CASE STUDIES OF QUESTIONS

Two case studies of mathematisation in PISA questions.

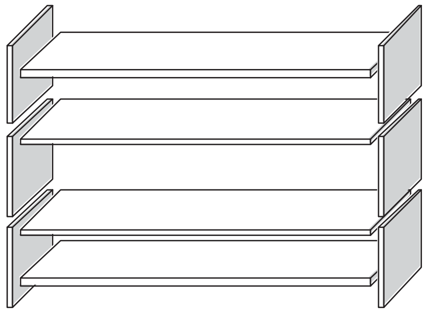
There is some real-world mathematical problem-solving present in all PISA mathematics questions. However, not all of the PISA mathematics questions make the full cycle of problem-solving clearly visible due to the limited time that students have to answer the questions: the average allowable response time for each question is around two minutes, which is too short a period of time for students to go through the whole problem-solving cycle. The PISA mathematics questions often require students to undertake only part of the problem-solving cycle and sometimes the whole problem-solving cycle. This section presents two case studies of questions where students are required to undertake the full problem-solving cycle.

The first case study: Bookshelves – Question 1

BOOKSHELVES

Question 1: BOOKSHELVES

To complete one set of bookshelves a carpenter needs the following components:



4 long wooden panels,

6 short wooden panels,

12 small clips,

2 large clips and

14 screws.

The carpenter has in stock 26 long wooden panels,
33 short wooden panels, 200 small clips, 20 large clips and 510 screws.

How many sets of bookshelves can the carpenter make?

Answer: ....

BOOKSHELVES SCORING QUESTION 1

Full Credit

Code 1: 5

No Credit

Code 0: Other responses

Code 9: Missing

STEP 1

The problem starts in a real-world context, and actually this reality is authentic. The problem is presented in a rather "natural" way, that is to say, there is a limited amount of text with a functional visual underpinning. However, the question is somewhat less complex than most problems are in reality due to the fact that there is almost no irrelevant or redundant information given in the question. This is important in light of step 2 of the problem-solving cycle: where students need to organise the facts in a more or less mathematical way and identify the relevant mathematics.

STEP 2

Many students will take a moment to check how the text stating the required components for a set of bookshelves relates to the picture. They will probably find out that this is not of much help since the only additional information is about how the long wooden panels relate to the short wooden panels.

So the relevant numbers are: one set requiring 4, 6, 12, 2 and 14; we have available 26, 33, 200, 20 and 510. These are the main components of step 2.

STEP 3

The students now translate the problem to the mathematical world. The question "how many sets can be completed before running out of one of the necessary components?" can be reformulated into a mathematical problem in the following way. Students need to look for the highest multiple of the first set (4, 6, 12, 2, 14) that fits into the other set (26, 33, 200, 20, 510).

STEP 4

In this step the students solve the problem.

One possibility for students that gives a high degree of confidence is producing the following table:

(4	6	12	2	14)	FOR 1 set
(8	12	24	4	28)	FOR 2 sets
(12	18	36	6	42)	FOR 3 sets
(16	24	48	8	56)	FOR 4 sets

Students list each row of components until they run out of one of the components:

(20	30	60	10	70)	FOR 5 sets
(24	36	72	12	84)	FOR 6 sets

Finally students run out of one component; there are only 33 short wooden panels available, and the last row to make a sixth set shows the need for 36 short wooden panels.

So, mathematically speaking, the highest multiple of

(4 6 12 2 14) that fits into (26 33 200 20 510) is 5.

It is very likely that students would use this strategy, but other strategies are also possible. Another possible strategy is to first identify the crucial component. If students really understand the problem right away from a more mathematical point of view, they might be tempted to calculate the ratios of the

components: $24/4 = 6 +$ remainder; $33/6 = 5 +$ remainder; $200/12$; $20/2$; $210/14$ are abundant, because each of them is greater than 10, so the answer is 5.

For a fully correct answer, students simply need to answer "5". Such an answer allows no further insights into the processes of mathematical problem solving followed by students. The correct answer is reached without concluding the problem-solving cycle.

STEP 5

To complete the problem-solving cycle, students would need to make sense of the meaning of the solution ("5") in terms of the real world. That is quite obvious here: with the listed available components, only five complete sets of bookshelves can be made. However, it is also possible to identify that the critical component is the short wooden panels, and that with three more of these it is possible to produce six sets.

Reflection on Bookshelves – Question 1

Although this problem seems straightforward, the difficulties involved in solving it should not be underestimated. There is a particular risk that lower achieving students could skip step 5, skip the reflection on the answer, and give 6 as an answer. These students would most likely use the "ratio" strategy, find the ratio to be 5.5 (6 to 33) and not reflect properly on the meaning of this number.

Double-digit coding could also be used to collect data on the different strategies students use, and the results could answer questions like:

- did the students use the first strategy: "build until you run out"?
- did the students use the "ratio" or "most critical component" strategy?
- did the students use another strategy?

The use of double-digit coding and specific answers and data associated with questions such as these would give us a better understanding about the nature and level of *mathematical literacy* of the students.

BOOKSHELVES Q1 is an example of a PISA mathematics question that requires a rather simple mathematical problem-solving process because students seem to know quite well what the problem is all about and how to solve it in a mathematical way. However, the PISA mathematics assessment also includes questions where the mathematical problem solving is more challenging, as there is no known strategy available to the students. SKATEBOARD Q3 is a good example of such a question.

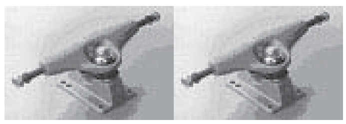
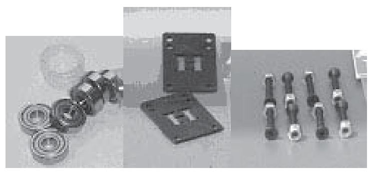
The second case study: Skateboard – Question 3

SKATEBOARD

Eric is a great skateboard fan. He visits a shop named SKATERS to check some prices.

At this shop you can buy a complete board. Or you can buy a deck, a set of 4 wheels, a set of 2 trucks and a set of hardware, and assemble your own board.

The prices for the shop's products are:

Product	**Price in zeds**	
Complete skateboard	82 or 84	
Deck	40, 60 or 65	
One set of 4 Wheels	14 or 36	
One set of 2 Trucks	16	
One set of hardware (bearings, rubber pads, bolts and nuts)	10 or 20	

Question 3: SKATEBOARD

Eric has 120 zeds to spend and wants to buy the most expensive skateboard he can afford.

How much money can Eric afford to spend on each of the 4 parts? Put your answer in the table below.

Part	**Amount (zeds)**
Deck	
Wheels	
Trucks	
Hardware	

SKATEBOARD SCORING QUESTION 3

Full Credit

Code 1: 65 zeds on a deck, 14 on wheels, 16 on trucks and 20 on hardware.

No Credit

Code 0: Other responses

Code 9: Missing

SKATEBOARD Q3 seems, at least at first glance, to have some similarities with BOOKSHELVES Q1. Students have to construct something, there are components, and both questions are presented in an authentic context. But mathematically speaking these questions are different, as the discussion will show.

STEP 1

The problem starts in a real-world context, and actually reflects a reality for many students in their daily life. For students who are unfamiliar with skateboards, photos are provided to give them some necessary information. SKATEBOARD Q3 is an example of a real situation for students as well. Students have a certain amount of money to spend and want to buy the best quality skateboard for their money.

STEP 2

It seems relatively straightforward for students to organise the problem. There are four components, and for three of the four, students need to make a choice (the only component for which there is no choice is the trucks). It is easy for students to identify the relevant mathematics since they have to add numbers and compare a sum with a given number.

A worksheet could look like:

Deck	40	60	65
4 Wheels	14	36	
Trucks	16		
Hardware	10	20	
TOTAL	120		

STEP 3

Mathematically speaking students have to find one number from each of the four categories that will result in the maximum sum within given restrictions. The restrictions for those numbers are: the first number has to be 40, 60 or 65; the second has to be 14 or 36; the third is 16; the fourth is either 10 or 20; and the sum cannot exceed 120. These are all the necessary elements to solve the problem.

STEP 4

Solving the mathematical problem is a little bit different than in BOOKSHELVES Q1 as there are no known strategies available to the students. This means that students will likely "fall back" on the trial-and-error method. This is actually a well known strategy, but every time it is applied, it is new within the context of the problem. Given the small amount of numbers that the students have to deal with, they can actually start making a list of all possibilities without running out of time. Given the task "to buy the most expensive", it seems advisable for students to start with the larger numbers from each collection:

65 + 36 + 16 + 20. These add up to 137, which is too much.

So students have to save 17 zeds. There are the following possibilities to save money:

On the deck:	5 or 25 zed
On the wheels:	22 zed
On the trucks:	nothing
On the hardware:	10 zed

This list makes the solution clear: save on the wheels, and students spend 115 zeds.

This strategy is structured. The problem with trial-and-error strategies lies often in the unstructured approach that students use. Students give different answers including:

40, 36, 16, 20

60, 14, 16, 20

60, 36, 16, 10

65, 36, 16, 20

The fact that students are not asked to give an explanation means that it is not possible to analyse their reasoning in more detail. A more detailed coding scheme like double-digit coding would allow for further insights into the use of actual strategies or reasoning and thinking.

STEP 5

This step was not tested in this question. It would be possible if students had been asked to explain their solutions. However, this question required students to fill out the numbers in a table. With appropriate argumentation, one of the solutions given above (40, 36, 16, 20) might be considered as a "better"

solution. For example, the student who came up with this answer might say that having excellent quality wheels is much more important than having a better deck. This might be a very good argument, indeed, but without the argument it is impossible to know whether this was actually the student's reasoning.

Reflection on Skateboard – Question 3

The problem-solving cycle becomes apparent in almost all aspects in SKATEBOARD Q3. The problem-solving strategy most often used is not a routine procedure. However, it is not possible to shed light on the actual problem-solving process, because in the present format, and with the restrictions of many large-scale tests, the relevant information is not collected for identifying the thinking and argumentation processes used to solve these problems. If such a question were used in daily practices of instruction in schools, it would offer opportunities for discussion and argumentation. It is possible to ask additional questions, and in particular, to require students to give arguments for their solutions.

STUDENTS' MATHEMATICAL UNDERSTANDINGS AND ITEM SCORING

Students' understanding of a topic varies across and within countries

This section will analyse students' understandings of particular mathematics topic areas overall and by subgroups. The section begins by surveying the nature of the information available from the different types of PISA mathematics questions, offering examples of information about students' mathematical understandings. Three more case studies of PISA 2003 mathematics items are then presented. The first examines results from a collection of questions that relate to proportional reasoning, which is a very important core mathematical topic. The second looks at those questions that involve some symbolic algebra and the third looks at average (mean). It is clear that form does make a difference when referring to the item format in which questions are presented (Braswell and Kupin, 1993; Traub, 1993; Dossey, Mullis, and Jones, 1993).

Item coding in the database and information on students' thinking

The data available from each question are dependent on the type of coding used to identify responses to the question. PISA has a variety of question formats, and each of them have been coded in a certain way (Table 6.1). For some questions, the students' answers were entered directly into software (*e.g.* the distractor, circled by students in the multiple-choice question or a simple numeric answer in some of the questions requiring short answers). Sometimes for technical reasons these responses were later recoded (*e.g.* numeric items were recoded automatically as 1 if the answer was correct and 0 otherwise). For PISA 2003, both recoded (scored) and actual (raw) information on such questions is available on the international database.

Other questions were coded by qualified coders following the international coding guides. Some of these questions can only provide data on the number of correct and incorrect responses or auxiliary information. The auxiliary information covers missing and invalid responses and the number of students to whom the question was not administered. Others (so called double-digit questions) provide possible insights into students' strategies, errors, cognitive obstacles and misconceptions.

Table 6.1
Use of different types of PISA 2003 mathematics question formats

Coding type	Number of questions
Directly entered responses	
Multiple-choice questions	17
Complex multiple-choice questions	11
Numeric response	21
Coded Responses	
Single-digit (including partial credit)	27(2)
Double-digit (including partial credit)	9(7)

In this section, examples of some of the coding types are given, along with some discussion of how this can be used to identify students' thinking.

Simple scoring of correct or incorrect answers (single-digit full credit coding)

Some questions only provide data on whether responses are correct, incorrect or missing. For example, the response categories for EXPORTS Q1 were recorded only as correct (Code 1, OECD average of 74%), incorrect (Code 0, OECD average of 17%) or missing (Code 9, OECD average of 9%).

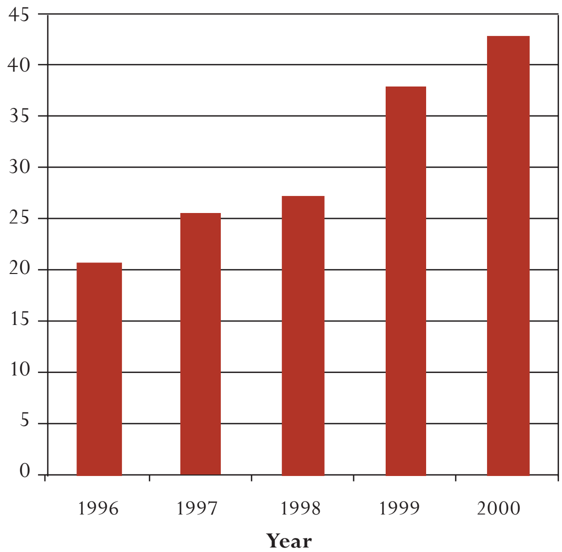
The coding of PISA questions provides clues to the student's understanding.

With this coding, the difficulty of these questions can be compared with that of other questions, both in terms of the percentage of students giving an incorrect answer and the percentage of students who do not attempt to answer the question. However, no information is available on the methods, correct or incorrect, used by students. The PISA study has adopted strict protocols to decide which constructed responses should be regarded as correct. In the case of the very easy EXPORTS Q1, students need to identify the height of the column (27.1) that is associated with 1998. Responses were marked correct if they answered 27.1 (without a unit), 27.1 million zeds, or 27 100 000 zeds. Even with such a simple question and simple response type, the criteria for correctness can make an important difference in student success.

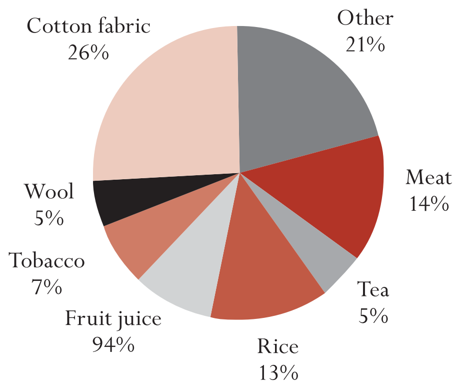
EXPORTS

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

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



Distribution of exports from Zedland in 2000



Question 1: EXPORTS

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

Answer: ....

EXPORTS SCORING QUESTION 1

Full Credit

Code 1: 27.1 million zeds or 27 100 000 zeds or 27.1 (unit not required)

No Credit

Code 0: Other responses

Code 9: Missing

Multiple-choice questions

Multiple-choice questions often provide more information, because the methods that students are likely to have used can sometimes be inferred from the choices (correct and incorrect) that were made. So, for example, EXPORTS Q2 is a multiple-choice question. Table 6.3 lists other examples.

Question 2: EXPORTS

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

A 1.8 million zeds.

B 2.3 million zeds.

C 2.4 million zeds.

D 3.4 million zeds.

E 3.8 million zeds.

EXPORTS SCORING QUESTION 2

Full Credit

Code 1: E. 3.8 million zeds

No Credit

Code 0: Other responses

Code 9: Missing

The distractors for EXPORTS Q2 are all calculated as approximately 9% of a quantity indicated on the bar graph (the total annual exports from 1996 to 2000 *i.e.* 9% of 20.4, 9% of 25.4, etc.).

Distractor C is the most frequent response after the correct response (E) both for the OECD average (Table 6.2) and also for a very large majority of countries. This is probably because the year 1998 was involved in the previous question EXPORTS Q1. This shows how student performance on a question can often be affected by irrelevant aspects of the question, the type or presentation of the question, or the students' failure to read all of the required information prior to answering.

These particular multiple-choice distractors have all been constructed in the same way, as 9% of a quantity on the column graph. This assumes that students find from the pie graph that 9% is associated with fruit juice, know that they need to find 9% of a quantity from the column graph, can calculate 9% of the quantity, but have difficulty identifying the correct quantity from the column graph.

Table 6.2
Distribution of responses for Exports – Question 2

	Distractor	Percent of students selecting the distractor (OECD average)	How the answer can be calculated
A	1.8 million zeds.	11%	9% of 1996 data = 1.836
B	2.3 million zeds.	10%	9% of 1997 data = 2.286
C	2.4 million zeds.	16%	9% of 1998 data = 2.439
D	3.4 million zeds.	8%	9% of 1999 data = 3.411
E	3.8 million zeds.	48%	9% of 2000 data = 3.834
	Missing response	7%	

Of itself, the coding provides no information about whether this is indeed an accurate description of the solution paths used by a majority of students. Students may have estimated 9% by eye (correctly or incorrectly), or they may have selected something a little less than 10% of 42.6 million (correctly or incorrectly).

Where students have not used a correct calculator or paper-and-pencil algorithm for finding the percentage, it is likely that they have calculated an answer not given amongst the multiple-choice distractors. For example, a common error in finding 9% would be to divide by 9, rather than to multiply, and this would lead to an answer greater than any of the distractors. Other potential obstacles for students that are not probed by the set of distractors are related to the use of millions. The problem would be easier for students who work directly in millions of zeds at every stage than for those students who convert to 42 600 000 zeds (or an incorrect version) and then have to convert back to the given form of the answer. This is likely to result in order of magnitude errors, which are also not tapped by these distractors. Students making calculation errors that result in answers not amongst the supplied distractors might omit the question, select the nearest, or try the calculation again.

Table 6.3
Examples of multiple-choice questions in Chapter 3

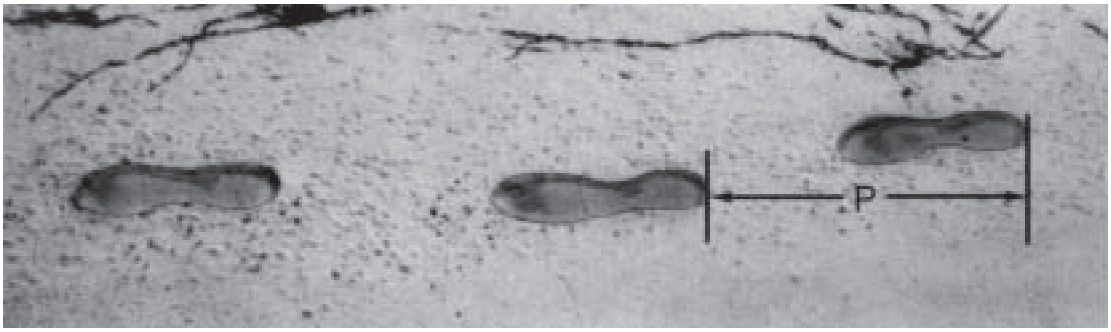
Question	Where to find question in Chapter 3
EXPORTS – Question 2	*Examples of easy questions section*
SKATEBOARD – Question 2	*Examples of questions of moderate difficulty section*
EARTHQUAKE – Question 1	
COLOURED CANDIES – Question 1	
NUMBER CUBES – Question 2[1]	
CARPENTER – Question 1[1]	*Examples of difficult questions section*

1. These questions are classified as complex multiple-choice questions as they require students to provide a series of correct answers from predefined choices.

Simple scoring of correct, incorrect or partially correct answers (single digit partial credit coding)

WALKING Q1 provides an example of a different coding pattern for responses called partial credit coding.

WALKING



The picture shows the footprints of a man walking. The pacelength P is the distance between the rear of two consecutive footprints.

For men, the formula, $\frac{n}{P} = 140$, *gives an approximate relationship between n and P where*

n = number of steps per minute, and

P = pacelength in metres.

Question 1: WALKING

If the formula applies to Heiko's walking and Heiko takes 70 steps per minute, what is Heiko's pacelength?

Show your work.

WALKING SCORING QUESTION 1

Full Credit

Code 2: 0.5 m or 50 cm, $\frac{1}{2}$ (unit not required)

$\frac{70}{P}$

70 = 140 P.

P = 0.5.

$\frac{70}{140}$

Partial Credit

Code 1: Correct substitution of numbers in the formula, but incorrect answer, or no answer.

$\frac{70}{P} = 140$ [substitute numbers in the formula only].

$\frac{70}{P} = 140$.

$70 = 140$ P.

P = 2 [correct substitution, but working out is incorrect].

OR

Correctly manipulated the formula into $P = \frac{n}{140}$, but no further correct work.

No Credit

Code 0: Other responses
70 cm.

Code 9: Missing

A correct response to this question is obtained by substituting the number of paces per minute ($n = 70$) into the formula $n/P = 140$ and finding P, the length of a pace. The maximum score of 2 is given to responses of 0.5 m, 50 cm or ½ m, with or without units. The partial credit score of 1 identifies a group of students with various incomplete algebra understandings: those who can substitute $n = 70$ into the formula, but do not solve the resulting equation correctly for P; those who can rearrange the formula to make the unknown pace length (P) the subject without going further; and those who obtain the likely incorrect answer $P = 2$. Other responses score 0. The OECD average on this question is 36% for the score 2, 22% for the score 1 and, 21% for the score 0 with 21% missing responses. In the PISA protocols of 2003, partial credit coding generally provides information on how much progress students have made towards a solution, rather than the type of progress that they have made.

Complex scoring recording the type of methods used in correct and partially correct answers (double-digit coding)

WALKING Q3 provides an example of "double-digit" coding (see Table 6.4 for further examples and the Chapter 3 section on examples of difficult questions in the PISA 2003 mathematics assessment). As with WALKING Q1 partial credit can be awarded in conjunction with "double-digit" coding. The double digits provide more information about the student's response and potentially about the method that they used and errors that they made (see Dossey, Jones and Martin, 2002).

Question 3: WALKING

Bernard knows his pacelength is 0.80 metres. The formula applies to Bernard's walking.

Calculate Bernard's walking speed in metres per minute and in kilometres per hour. Show your working out.

WALKING SCORING QUESTION 3

Full Credit

Code 31: Correct answers (unit not required) for both metres/minute and km/hour:
N = 140 × 0.80 = 112.
Per minute he walks 112 × 0.80 metres = 89.6 metres.
His speed is 89.6 metres per minute.
So his speed is 5.38 or 5.4 km/hr.

Code 31 as long as both correct answers are given (89.6 and 5.4), whether working out is shown or not. Note that errors due to rounding are acceptable. For example, 90 metres per minutes and 5.3 km/hr (89 × 6) are acceptable.
89.6, 5.4.
90, 5,376 km/h.
89.8, 5376 m/hour [note that if the second answer is given without units, it should be coded as 22].

Partial Credit (2 points)

Code 21: As for Code 31 but fails to multiply by 0.80 to convert from steps per minute to metres per minute. For example, his speed is 112 metres per minute and 6.72 km/hr.
112, 6.72 km/h.

Code 22: The speed in metres per minute correct (89.6 metres per minute) but conversion to kilometres per hour incorrect or missing.
89.6 metres/minute, 8960 km/hr.
89.6, 5376.
89.6, 53.76.
89.6, 0.087 km/h.
89.6, 1.49 km/h.

Code 23: Correct method (explicitly shown) with minor calculation error(s) not covered by Code 21 and Code 22. No answers correct.
n = 140 × 0.8 = 1120; 1120 × 0.8 = 896. He walks 896 m/min, 53.76 km/h.
n = 140 × 0.8 = 116; 116 × 0.8 = 92.8. 92.8 m/min → 5.57 km/h.

Code 24: Only 5.4 km/hr is given, but not 89.6 metres/minute (intermediate calculations not shown.

WALKING Q3 is complex. From the given value of *P*, the pace length, students have to use the formula to work out *n*, how many paces per minute, multiply *P* by *n* to get the distance travelled per minute and then convert this to kilometres per hour. As partially correct answers are credited for this question, there are several possible scores: 3 for a fully correct answer; 2 for a high-level partially correct answer; 1 for a low-level partially correct answer, or 0 for an incorrect answer. However, information is also separately recorded about the methods used by students and other features of their solutions. In this case, the following errors are separately coded for students scoring 2 points:

- finding *n* correctly (steps per minute) but not multiplying by *P* to get metres per minute, but working correctly in all other respects (code 21);
- incorrect conversions of the correct metres per minute speed to km/hr (code 22);
- minor calculation errors in a fully correct method (code 23);
- correct km/hr speed without supporting calculations (code 24).

WALKING Q3 was a difficult question. The OECD average of students scoring 3 (fully correct) was 8% and the total of students falling into all codes of the high-level partially correct answer (score 2) was 9%. Across the OECD countries, the most common reasons why students did not get the answer fully correct (score 3) were a failure to calculate metres per minute (4.8% of students on average) or an incorrect conversion to kilometres per hour (3.3% of students on average). Only a minority of students made minor calculations errors (0.58% of students on average) or gave the correct kilometres per hour speed without supporting calculations (0.22% of students on average).

The countries with high percentages of students giving fully correct answers to this question (score 3) also tend to have high percentages of students who just fall short of giving fully correct answers (score 2): in fact there is a strong correlation (0.86 across the OECD countries) between these two groups of students. For example, the only countries with more than 16% of students giving fully correct answers were Japan (18%) and the partner economy Hong Kong-China (19%). These countries, and other countries with relatively high percentages of students scoring 3, had the highest percentages of students scoring 2. Only in Japan (14%) and the partner economies Hong Kong-China (22%) and Macao-China (18%) were there more than 10% of students in code 21 for WALKING Q3 and only in the partner economies Hong Kong-China (4%) and Macao-China (4%) were there more than 0.5% of students in code 24. For code 22, the partner country Liechtenstein had the highest percentage of students (8%) and Liechtenstein had the fourth highest percentage of students scoring on this question overall (32% correct). Five to six percent of students scored in code 22 in the Czech Republic, the Slovak Republic, France, Poland, Hungary and the partner country the Russian Federation and this was the most common reason why students did not score fully correct answers in these countries.

Countries with similar performance patterns over the PISA mathematics assessment also show similarities in the frequency of double-digit codes. For example, it was noted in Chapter 3 that the partner economies Hong Kong-China and Macao-China have similar performance patterns across PISA mathematics questions. They also show similar patterns of responses within questions using the double-digit codes.

The aim of PISA is to measure *mathematical literacy*, therefore the use of double-digit coding generally indicates whether students have used a correct method to solve the mathematical problem, but have small calculation errors, rather than to show the nature of the errors that students make or to measure the incidence of established misconceptions and common error patterns.

Double coding can help disentangle student's problem-solving strategies and understanding.

The above summary of types of coding shows that each question format can provide interesting data for analysis in relation to students' approaches to mathematical problems. Simple multiple-choice questions with one correct answer among a set of well-constructed distractors can provide valuable information about the prevalence of misconceptions, at least as they are restricted to students' selecting them from a list of alternatives. Items calling for students to construct and provide a simple numeric response provide slightly better information in that students have to show what they would do without being prompted by a set of distractors. The analysis still provides correctness and distribution data.

At yet a higher level, single-digit coded items with partial credit offer a deeper insight into students' performances. In many countries, such partial credit scoring shows that many students start correctly, move to an intermediate result, and then fail to complete the task set by the problem in the unit. Partial credit shows that they have command of subsidiary knowledge and skill, even while students did not attend to the full task presented. Finally, double-digit coding provides mathematics educators with a picture of the degree of correctness, including partial credit, as well as a picture of the relative distribution of strategies employed by students. All coding types provide information about difficulty of items, but additional information shows the differing depths of understanding that exist within and between countries.

Table 6.4
Examples of questions with double-digit coding in Chapter 3

Question	Where to find question in Chapter 3
GROWING UP – Question 2	*Examples of the easiest questions section*
GROWING UP – Question 3 EXCHANGE RATE – Question 3	*Examples of questions of moderate difficulty section*
WALKING – Question 3 ROBBERIES – Question 1	*Examples of difficult questions section*

STUDENTS' UNDERSTANDING OF PROPORTIONAL REASONING

A wide range of important mathematical ideas with strong application in the real world involve proportional reasoning. Proportional reasoning is an important part of the "conceptual field of multiplicative structures" (Vergnaud 1983, 1988), which consists of situations where multiplication and division are usually required, including problems involving ratios, rates, scales, unit conversions, direct variation, fractions, and percentages as well as simple multiplication and division problems involving discrete objects or measures. It takes a very long time for students to master multiplicative structures. Additive structures (involving addition and subtraction) are principally developed in the early grades, but understanding of multiplicative structures continues to develop from the early years of schooling through to the PISA age groups. A critical part of the learning in central content areas is the differentiated roles played by conceptual content and procedural content and the building of connections between "knowing what" and "knowing how" (Hiebert, 1986).

The prevalence of proportional reasoning in PISA questions

Evidence of the importance of proportional reasoning to a mathematically literate person comes from the observation that 11 of the 85 questions in the PISA 2003 mathematics assessment involved proportional reasoning, and this contributed to many others. For example, EXPORTS Q2 requires knowledge of percentages, and the unit EXCHANGE RATE requires conversion of currencies. Table 6.5 gives some other examples. Ability in proportional reasoning underlies success in many

Table 6.5
Instances of proportional reasoning in questions presented in Chapter 3

Unit name	Questions (see Chapter 3)	Skills required	Percent correct (OECD average)
WALKING	Question 3	Includes conversion of metres per minute to kilometres per hour (minor part of question) with other skills.	8%
GROWING UP	Question 3	Involves concept of rate of growth, with other skills.	45%
ROBBERIES	Question 1	An absolute difference is to be judged as a relative difference.	14%
EXCHANGE RATE		Converting currencies by	
	Question 1	multiplying by 4.2	80%
	Question 2	dividing by 4.0 and	74%
	Question 3	explaining which rate is better.	40%
EXPORTS	Question 2	Finding 9% of 42.6 million, along with graph reading skills.	46%
COLOURED CANDIES	Question 1	Finding a probability (relative frequency), with simple graph reading skills.	50%
STAIRCASE	Question 1	Apportioning a rise of 252 cm over 14 stairs.	78%

aspects of the school curriculum so teachers need to pay particular attention to students' understandings of this important area. There are underlying proportional reasoning abilities, as well as specific knowledge required to deal with its many occurrences (such as probability, percentage, speed, slope, rate of change, *etc.*).

The difficulty of proportional reasoning questions

Within the mathematics education literature, major factors that contribute to the difficulty of proportional reasoning have been identified. Some important factors are the nature of the real context and the numerical nature of the ratio, rate, or proportion involved. For example, students deal with simple ratios much more easily, both procedurally and conceptually, than with complex ratios. It is much easier to identify that a problem can be solved by multiplying a quantity by 3 than that it can be solved by multiplying by $^{11}/_{12}$. When the rate, ratio or proportion is a number between 0 and 1, students often choose division when multiplication is appropriate, and vice versa.

One simple and useful classification of levels of proportional reasoning questions is provided by Hart (1981). The Hart scale provides a guide to classifying the difficulty of proportional reasoning questions. The description of the Hart levels is given in Table 6.6 and selected PISA questions that have a major emphasis on proportional reasoning, with question difficulties and PISA proficiency levels are given in Table 6.7.

The data in Table 6.7 indicate that the proportional reasoning questions show a wide range of difficulty, from -1.85 to 3.21, although the difficulty of the hardest item, the unreleased THE THERMOMETER CRICKET Q2, is increased by the use of algebra. The Hart scale and the PISA scale basically follow a similar pattern with two exceptions (EXCHANGE RATE Q3 and POPULATION PYRAMIDS Q3 to be discussed below). Classification of PISA questions testing proportional reasoning according to the Hart (1981) scale shows that the majority of these questions could be classified as having a difficulty of two or three. Hart considered that students did not really display proportional reasoning until they were able to be successful on level 3 items. Since the PISA difficulty of the Hart level 2 items is around 0, this indicates that the development of proportional reasoning ability must still be considered an important instructional goal for teachers of many 15-year-old students.

Table 6.6
Hierarchy of proportional reasoning items (Hart, 1981)

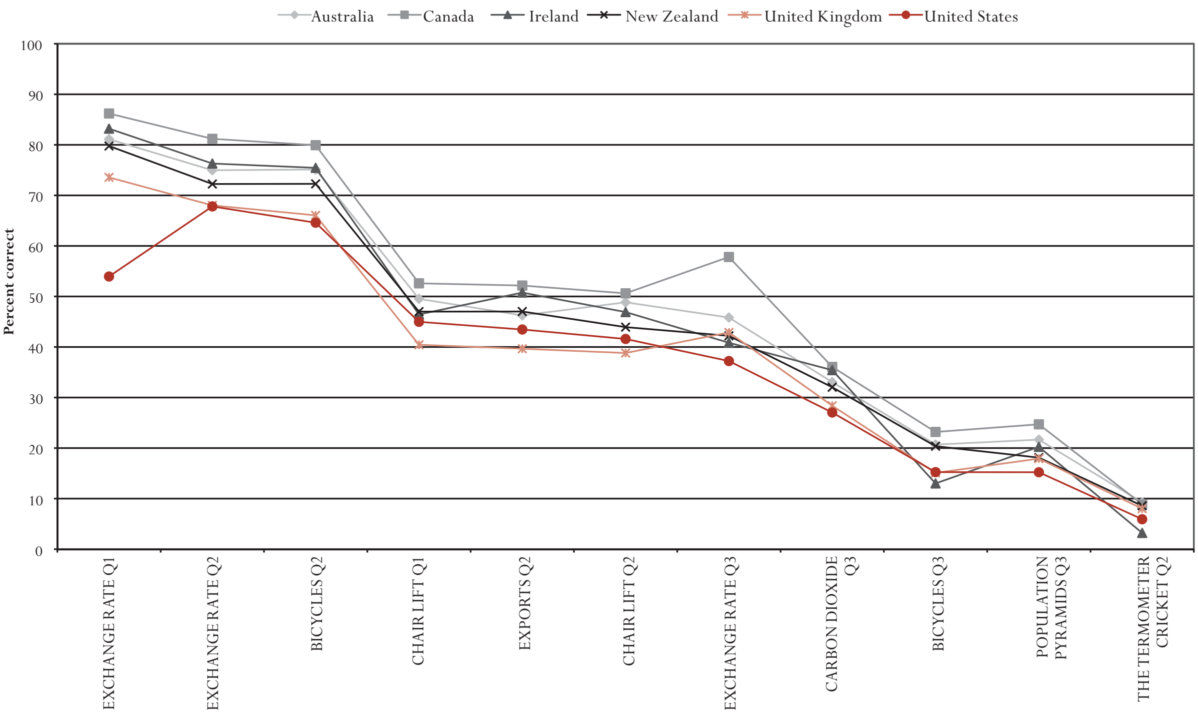
Level	Description and examples
1	Problems involving doubling, trebling or halving. Rate given.
2	Rate easy to find or given. Also, problems which can be solved by adding appropriate quantity and half the quantity. Example: to change a recipe for 4 to a recipe for 6, take the original recipe and add half of each quantity.
3	Rate is more difficult to find. Also, when fraction operations are involved. Example: Mr Short is 4 matchsticks tall or 6 paperclips. Mr Tall is 6 matchsticks tall; how tall in paperclips?
4	Use of ratio or rate needs to be identified. Questions complex in numbers in ratio or in setting. Example: making an enlargement in the ratio of 5:3.

Table 6.7
Level of difficulty of proportional reasoning questions
(PISA proficiency level, question difficulty parameter, Hart level)

PISA question	PISA proficiency level	Question difficulty parameter	Hart level
EXCHANGE RATE – Question 1	1	-1.85	1
EXCHANGE RATE – Question 2	2	-1.36	2
BICYCLES – Question 2	2	-1.20	2
CHAIR LIFT – Question 1	4	0.04	2
EXPORTS – Question 2	4	0.14	3
CHAIR LIFT – Question 2	4	0.22	3
EXCHANGE RATE – Question 3	4	0.45	2
CARBON DIOXIDE – Question 3	5	0.93	3
BICYCLES – Question 3	5	1.55	4
POPULATION PYRAMIDS – Question 3	6	1.71	3
THE THERMOMETER CRICKET – Question 2	6	3.21	4

The results of the PISA proportional reasoning questions showed wide variation across countries, but groups of countries which have been shown to perform similarly in other studies once again performed similarly. For example, Figure 6.2 shows that the pattern of performance of a selection of the English-speaking countries is very similar. The items are arranged in rank order of question difficulty.

Figure 6.2 ■ **Performance of some English speaking countries on proportional reasoning items, illustrating their similar pattern of performance**



EXCHANGE RATE Q3, which is seen to behave unusually in Table 6.11, shows an interesting variation.

EXCHANGE RATE

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

Question 1: EXCHANGE RATE

Mei-Ling found out that the exchange rate between Singapore dollars and South African rand was:

1 SGD = 4.2 ZAR

Mei-Ling changed 3000 Singapore dollars into South African rand at this exchange rate.

How much money in South African rand did Mei-Ling get?

Answer: ....

EXCHANGE RATE SCORING QUESTION 1

Full Credit

Code 1: 12 600 ZAR (unit not required)

No Credit

Code 0: Other responses

Code 9: Missing

Question 2: EXCHANGE RATE

On returning to Singapore after 3 months, Mei-Ling had 3 900 ZAR left. She changed this back to Singapore dollars, noting that the exchange rate had changed to:

1 SGD = 4.0 ZAR

How much money in Singapore dollars did Mei-Ling get?

Answer: ....

EXCHANGE RATE SCORING QUESTION 2

Full Credit

Code 1: 975 SGD (unit not required)

No Credit

Code 0: Other responses

Code 9: Missing

Question 3: EXCHANGE RATE

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

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

EXCHANGE RATE SCORING QUESTION 3

Full Credit

Code 11: Yes, with adequate explanation.

- Yes, by the lower exchange rate (for 1 SGD) Mei-Ling will get more Singapore dollars for her South African rand.
- Yes, 4.2 ZAR for one dollar would have resulted in 929 ZAR. [Note: student wrote ZAR instead of SGD, but clearly the correct calculation and comparison have been carried out and this error can be ignored]
- Yes, because she received 4.2 ZAR for 1 SGD and now she has to pay only 4.0 ZAR to get 1 SGD.
- Yes, because it is 0.2 ZAR cheaper for every SGD.
- Yes, because when you divide by 4.2 the outcome is smaller than when you divide by 4.
- Yes, it was in her favour because if it didn't go down she would have got about $50 less.

No Credit

Code 01: Yes, with no explanation or with inadequate explanation.

- Yes, a lower exchange rate is better.
- Yes, it was in Mei-Ling's favour, because if the ZAR goes down, then she will have more money to exchange into SGD.
- Yes, it was in Mei-Ling's favour.

Code 02: Other responses.

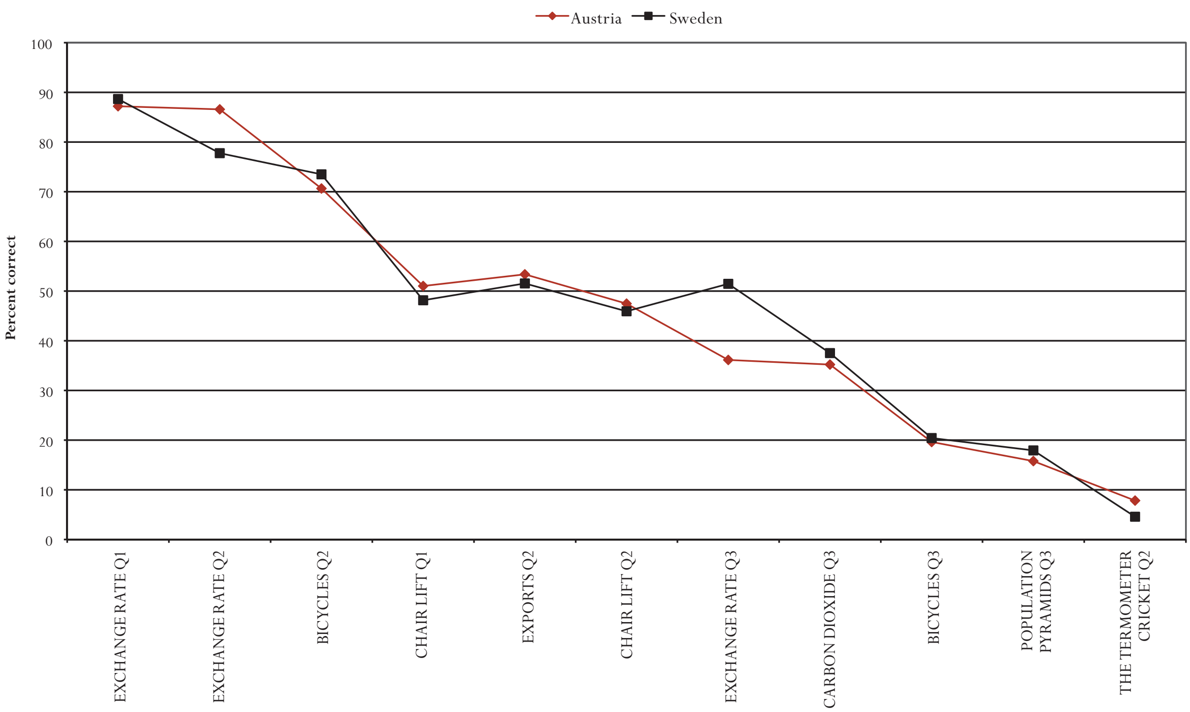
Code 99: Missing.

Figure 6.3 shows examples of two pairs of otherwise similar countries Austria and Sweden, performing differently on this question. Whereas the first two questions in the EXCHANGE RATE unit required calculations, the third question, probed understanding more conceptually (see Chapter 4).

The proportional reasoning questions show a small gender difference. On average, the percent correct for males is 2% greater than for females. This is independent of the question difficulty. This is an unexpected finding since it is often the case that gender differences are most pronounced with high performance.

These looks into proportional reasoning provide mathematics educators and curriculum specialists with comparative understanding of 15-year-olds' understanding of aspects of proportionality. The links to Hart's theoretical model

Figure 6.3 ■ **Proportional reasoning performances of Austria and Sweden, showing variation in Exchange Rate – Question 3**



and the correlation of the PISA difficulty findings with the hierarchical levels of the model indicates that PISA findings can be used as one data source for the possible validation of such theoretical models. But, beyond that, the data on proportionality serves as an international look at differences in the deeper understandings students have, country-by-country, in a mathematical area directly underlying the study of linear equations.

STUDENTS' UNDERSTANDING OF SYMBOLIC ALGEBRA

This section looks at student performance on the questions explicitly using symbolic algebra. There are very few such questions, which reflect the on-going debates in the mathematics education community about the place of symbolic algebra in the school curriculum. This topic has been greatly affected by the "massification" of schooling, and the lack of obvious relevance of algebraic symbols in students' everyday lives has meant that there has been substantial questioning of its role in many countries, often resulting in substantial adjustment to the algebra curriculum.

The PISA questions succeed in finding contexts and problems that are properly part of *mathematical literacy*, and show that even symbolic algebra has a place in *mathematical literacy*. At the same time, the low number of such questions within the PISA 2003 mathematics assessment indicates that dealing with algebraic symbols is not of extremely high importance for *mathematical literacy*

(MacGregor and Stacey, 1997; Stacey and MacGregor, 2000). Out of seven PISA questions classified as Algebra (see Annex A3) six questions use algebraic symbols explicitly, while the other uses only the graphical representation. This question is the non-released Q2 of the CONTAINERS unit.

In the PISA framework, the *change and relationships* overarching idea is rightly recognised to be much larger than the set of problems where algebraic letter symbols are used. Indeed it is also larger than algebra, as it includes work about patterns, functions and variation that can be represented graphically, numerically and spatially as well as symbolically.

Letter symbols appear in PISA in formulas, which is one of their several uses within and outside mathematics. Letters in a formula are generally relatively easy to conceptualise because they have a clear referent (a quantity to be used in calculation). Moreover, the purpose of a formula is clear: to calculate one quantity from others. For example, in the formula $A = LW$, the letter A stands for the area of a rectangle, and L and W stand for side lengths. The purpose of the formula is to calculate the area (A) from its length (L) and width (W). Happily, the characteristics that make formulas a relatively easy part of symbolic algebra also make them a very important part of early algebra.

The graphs in Figure 6.4 and Figure 6.5 show the performance of selected countries on the six questions that used letter symbols in a formula. The unit WALKING contains two of them (see Chapter 3). OECD average percent

Figure 6.4 ■ **Performance on algebra items for countries scoring highly on the content items from *change and relationships***

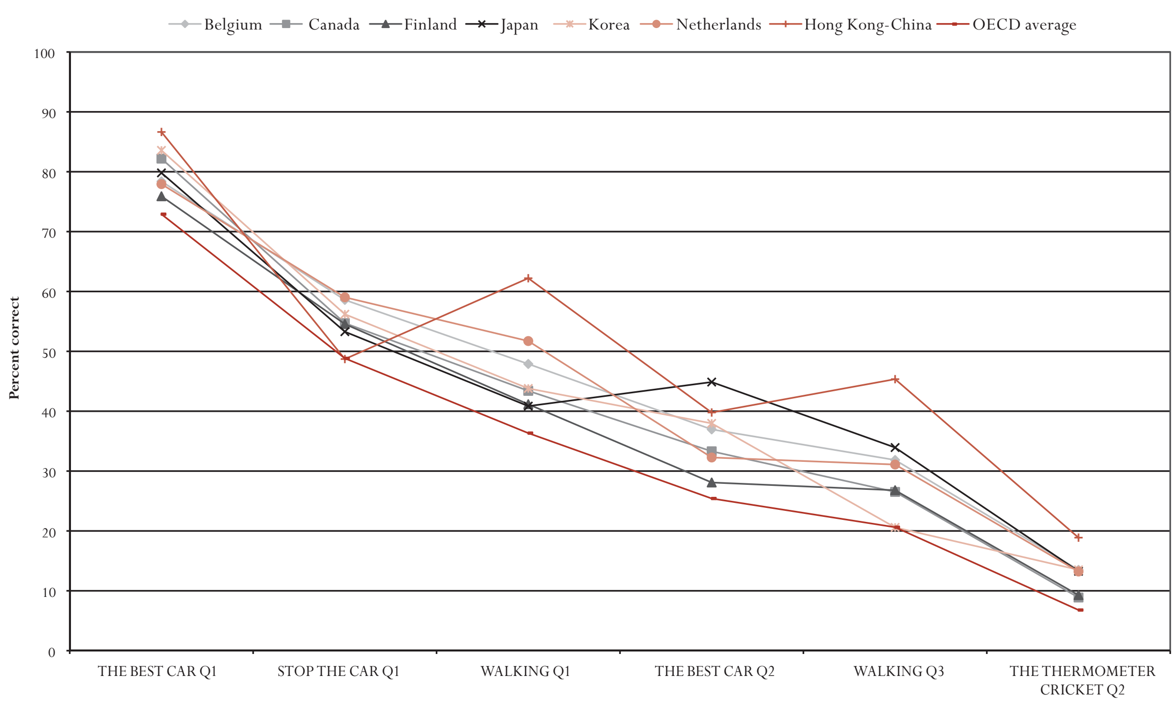
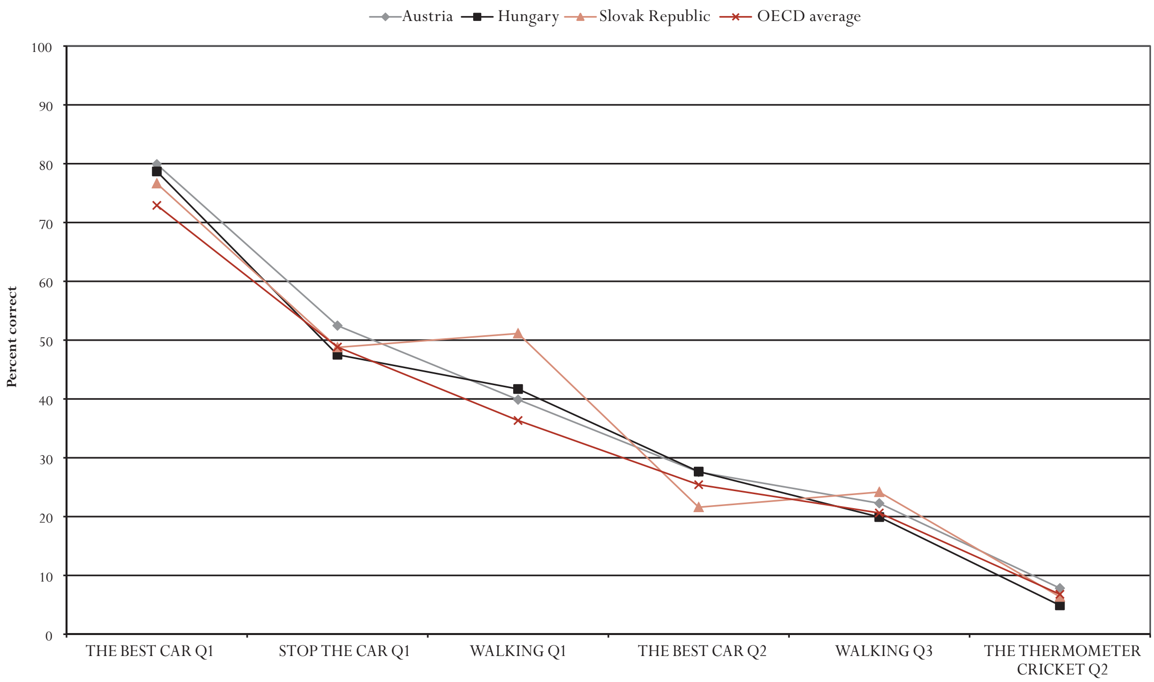


Figure 6.5 ■ **Performance on algebra items for the countries scoring at the OECD average on the content items from *change and relationships***



correct is 35% for WALKING Q1 and 20% for the more difficult WALKING Q3. Figure 6.4 shows the performance of the best performing countries in the *change and relationships* overarching area and Figure 6.5 shows the performance of a group of countries not statistically different from the OECD average.

These six questions show a wide range of performance. Examination of the use of algebraic symbols in each question shows little link to the algebraic demand of the questions (substituting, interpreting, using and writing formulas). For example, the most complicated algebraic expressions are in the second easiest question (non-released question STOP THE CAR Q1). However, it is the case that writing a formula from worded information was required only in the most difficult (non-released question THE THERMOMETER CRICKET Q2). As noted above, this question also had a high proportional reasoning demand, so the writing of the formula does not entirely explain the difficulty. The variation in performance on WALKING Q1 and WALKING Q3 (see Chapter 3), which is evident in both Figures 6.4 and 6.5, may indicate that students had difficulty interpreting this question, especially in some languages. Some evidence for this is that the western English-speaking countries do not change rank order on this question, as the countries in Figure 6.4 and Figure 6.5 do.

None of these six questions have double-digit coding that relates directly to students' understanding of algebra or the common algebraic errors.

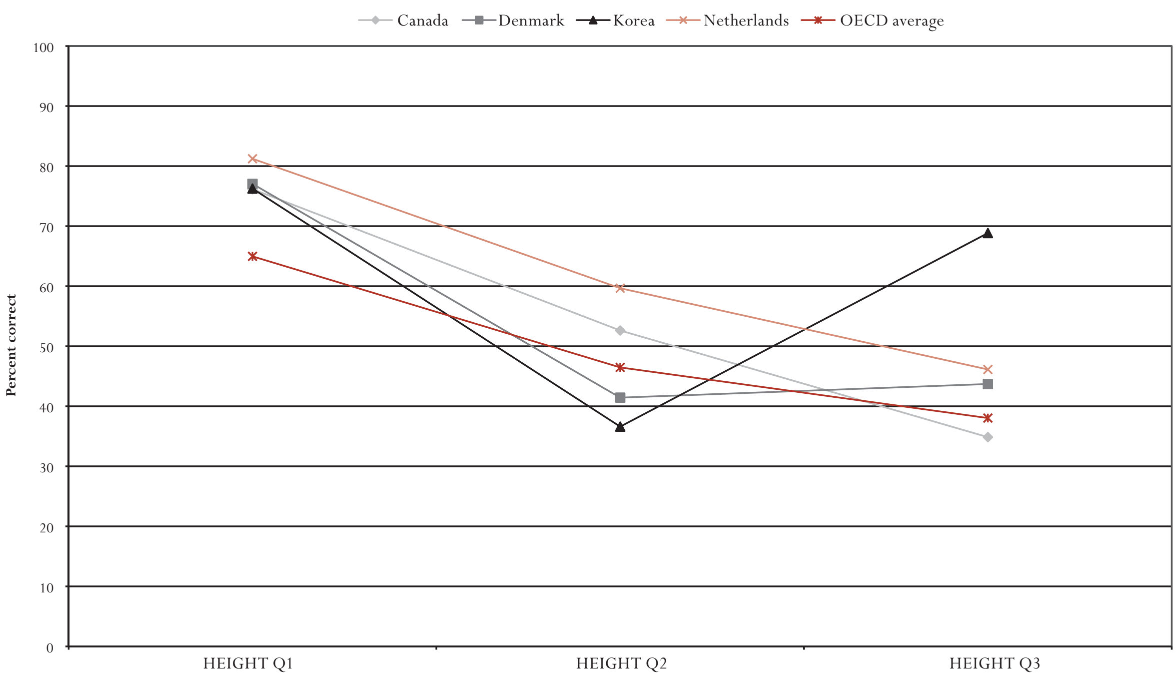
STUDENTS' UNDERSTANDING OF AVERAGE

The following Figure 6.6 and Figure 6.7 show the percent correct for selected high-performing countries and the OECD average, on several non-released items that involve different types of knowledge about the average of a data set.

Within the real-world context of the non-released PISA unit HEIGHT, HEIGHT Q1 required an explanation of the procedure of calculating an average, HEIGHT Q2 tested knowledge of some of the properties of average, and HEIGHT Q3 required a difficult calculation of average.

Figure 6.6 shows that these countries have similarly high success rates on the item that tests knowledge of how to calculate an average from data, all above the OECD average. In fact, the variability of the country averages on this item is considerably larger than on the other items, but the selection of countries hides this fact. The central item of the figure, HEIGHT Q2, shows that the percentage of students with good knowledge of the properties of averages from these countries is spread, with the Netherlands having the highest percent correct, and some countries that were high performing on the previous item having less than the OECD percent correct. The third item, HEIGHT Q3, requires a difficult calculation of average, which can be made easier by a good conceptual understanding, as has been tested in the previous item. Here the graph shows that Korea, with relatively poor conceptual understanding performs extremely well. This is an interesting result, which is also evident in the results of Japan, although the effect is less marked.

Figure 6.6 ■ **Results of selected countries on HEIGHT concerning the mathematical concept of average**[1]



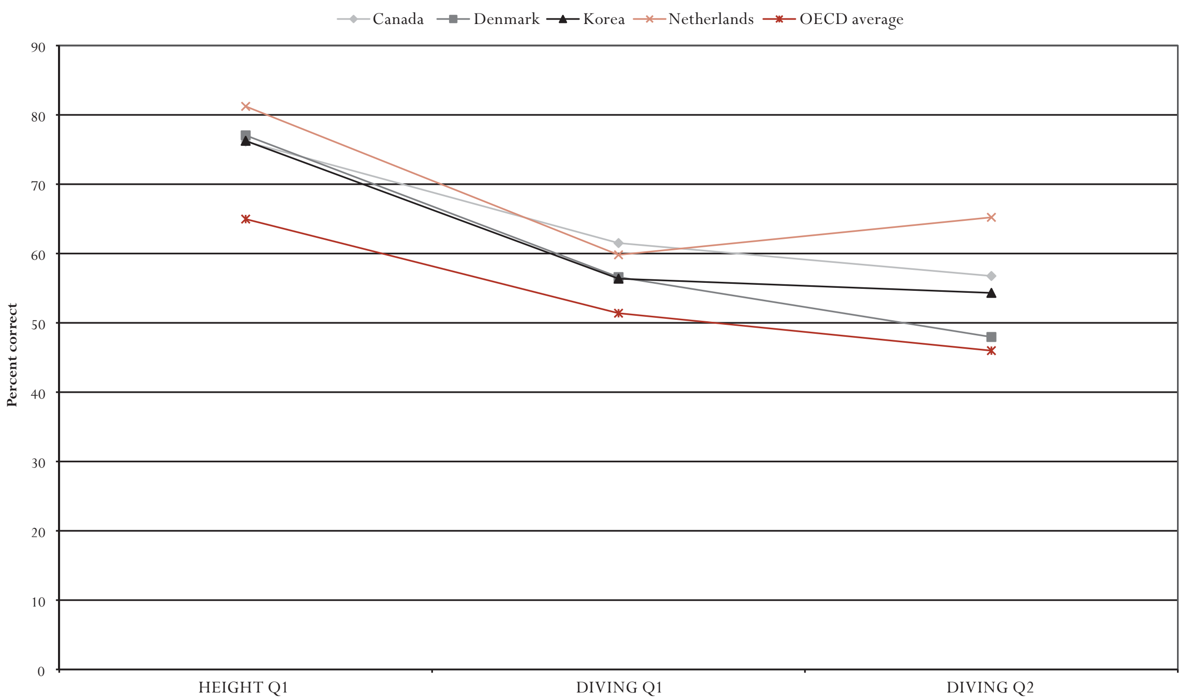
1. This is the percentage scoring 3 or 4 out of 4, on this partial credit item.

It seems that for some of the countries displayed, the calculation of the difficult average has been a challenging problem, where students have had to marshal their own mathematical problem solving resources. Good conceptual knowledge of average may have helped to solve this item. On the other hand, Korea shows a contrasting effect. Here there are the same percent correct on the difficult and straightforward items. This indicates that Korean students are able to treat this problem solving item as a routine task.

The three items in Figure 6.7 also show a progression. HEIGHT Q1, the first item in the graph (and the previous graph), requires the explanation for calculating an average. The selected countries nearly retain this advantage to the second item of Figure 6.7, which requires the calculation of an average embedded in a more complex situation. The additional complexity has caused an average drop of about 13% in percent correct. The third item again requires interpretation related to the concept of average and its calculation. The Netherlands and Canada again have a high percent correct, as they did for item HEIGHT Q2 in Figure 6.9 and Denmark again shows relatively less conceptual knowledge of the meaning of average. However, in this instance, just as many Korean students can interpret the real situation as can carry out the complicated calculation of average. The interpretation in this item tends towards interpreting an effect on a calculation, rather than linking average to the original data set.

This may indicate that in the Korea, instruction focuses less on the meaning of average in terms of the original data set and more on the attributes of the formula.

Figure 6.7 ■ **Results of selected countries on some non-released items concerning the mathematical concept of average**



CONCLUSION

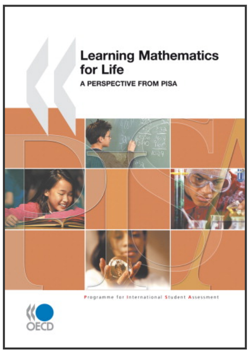
The PISA Framework of *mathematical literacy* (OECD, 2003) described the problem-solving process in terms of the process of "mathematisation". This chapter presented two examples of PISA items that make the problem-solving cycle visible. In these examples the problem-solving cycle comes alive in almost all aspects of the questions. Each problem solving strategy was not a routine procedure.

Unfortunately the authors cannot shed light on the actual specific strategy students used. The PISA scoring format does not provide specific information on the thinking and argumentation processes students actually used in solving problems.

At the level of daily practices of instruction in schools, however, it is possible to ask additional questions and, particularly, let students give arguments for their solutions. Teachers and other researchers might try using PISA items with their students and compare their results with those observed in this chapter.

Despite these limitations, the PISA database contains some insights on the problem-solving processes students use to tackle problems. First and foremost, the mathematical content and the selected contexts of the questions provide important examples of how mathematics is likely to be used in everyday life and work. Many items, for example, directly or indirectly involve proportional reasoning, which stresses the importance of this topic for *mathematical literacy*. The very large number of graphs in PISA questions also reflects the importance of this topic to modern living. Teachers might ask their students to relate problem structures presented in their classes to contexts from their personal lives to better connect the mathematics to students' self perceived needs and experiences.

When examining the PISA database through the lens of students' thinking about an individual mathematical topic, there is often little relevant data. The codes used in PISA aim principally to provide a good picture of mathematics in use. If there are good data for other purposes, it is serendipitous. Looking through the lens of a single mathematical topic, there may not be many PISA items to analyse. Furthermore, the coding may not record the common errors and any effect of a particular approach to the mathematical topic is likely to be masked by the mix of skills that is characteristic of mathematics being used in a context. In the future, to better reveal aspects of students' intra-mathematical thinking, it may be possible to sharpen the distractors used in multiple choice items and to include double digit coding. However, this should not be done at the expense of PISA primary goals.



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), “Mathematical Problem Solving and Differences in Students' Understanding”, in *Learning Mathematics for Life: A Perspective from PISA*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264075009-8-en

OECD