A Profile of Student Performance in Mathematics

INTRODUCTION

The PISA 2000 results raised issues about student performance both across and within countries…

Since 1997, OECD governments have collaborated to monitor the outcomes of education in terms of student performance on a regular basis and within an internationally agreed common framework. The first PISA assessment, carried out in 2000, revealed wide differences in the extent to which countries succeed in equipping young adults with knowledge and skills in reading, mathematics and science. For some countries, the results were disappointing, showing that their 15-year-olds' performance lagged considerably behind that of other countries (and perhaps their own expectations) sometimes by the equivalent of several years of schooling[1] and in certain cases despite high investments in education. PISA 2000 also highlighted significant variation in the performance of schools and raised concerns about equity in the distribution of learning opportunities.

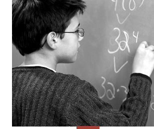
…and while the overall results in 2003 have changed only slightly, country differences continue to evolve.

Among the 25 OECD countries for which performance can be compared between 2000 and 2003, average mathematics performance increased in one of the two content areas measured in both surveys. For the other mathematical content area, as well as for science and reading, average performance among OECD countries has remained broadly unchanged. However, performance has changed in different ways across OECD countries. Finland, the top performing country in the PISA 2000 reading assessment, has maintained its high level of reading performance while improving its performance in mathematics and science.[2] This now places Finland on a par in mathematics and science with the previously unmatched East Asian countries. By contrast, in Mexico, the lowest performing OECD country in the 2000 assessment, the pressure to expand the still limited access to secondary education[3] may have been one of the factors putting strains on educational quality, with performance in the 2003 assessment lower in all three assessment areas.

This chapter reports results in mathematics, the main focus in PISA 2003…

This chapter presents in detail the results from the PISA 2003 mathematics assessment. Mathematics is the main focus of PISA 2003, and accounted for over half of all assessment time. This allowed mathematics performance to be assessed more thoroughly than in PISA 2000, and for its measurement to be refined.

- The chapter begins by setting the results in the context of how mathematics is defined, measured and reported. It considers a series of key questions. What is meant by "mathematical literacy"? In what ways is this different from other ways of thinking about mathematical knowledge and skills? Why is it useful to think of mathematical competencies in this way, and how can the results be interpreted?
- In the second part, the chapter examines student performance in mathematics. Since results vary in important ways across the four content areas of mathematics examined in PISA 2003, the analysis is described separately for each content area before a summary picture is presented at the end.
- In as much as it is important to take the socio-economic context of schools into account when comparing school performance, any comparison of the



outcomes of education systems needs to account for countries' economic circumstances and the resources that they devote to education. To address this, the third part of the chapter interprets the results within countries' economic and social contexts.

…while further chapters report other outcomes: student approaches to learning and performance in reading and science.

Chapter 3 continues the analysis of student outcomes by examining a wider range of student characteristics that relate to performance in mathematics and that can be considered important educational outcomes in their own right, including students' motivation to learn mathematics, their beliefs about themselves and their learning strategies in mathematics. Later, Chapter 6 extends the reporting of student outcomes in PISA 2003 by looking at performance in reading and science.

THE PISA APPROACH TO ASSESSING MATHEMATICS PERFORMANCE

How mathematics is defined

Today, all adults need a solid foundation in mathematics to meet their goals.

For much of the last century, the content of school mathematics and science curricula was dominated by the need to provide the foundations for the professional training of a small number of mathematicians, scientists and engineers. With the growing role of science, mathematics and technology in modern life, however, the objectives of personal fulfilment, employment and full participation in society increasingly require that all adults – not just those aspiring to a scientific career – be mathematically, scientifically and technologically literate.

PISA defines a form of mathematical literacy…

PISA therefore starts with a concept of mathematical literacy that is concerned with the capacity of students to analyse, reason and communicate effectively as they pose, solve and interpret mathematical problems in a variety of situations involving quantitative, spatial, probabilistic or other mathematical concepts. *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills* (OECD, 2003e) through which OECD countries established the guiding principles for comparing mathematics performance across countries in PISA, defines mathematical literacy as "…an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgements and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen" (OECD, 2003e).

…that requires engagement with mathematics…

When thinking about what mathematics might mean for individuals, one must consider both the extent to which they possess mathematical knowledge and understanding, and the extent to which they can activate their mathematical competencies to solve problems they encounter in life. PISA therefore presents students with problems mainly set in real-world situations. These are crafted in such a way that aspects of mathematics would be of genuine benefit in solving the problem. The objective of the PISA assessment is to obtain measures of the extent to which students presented with these problems can activate their mathematical knowledge and competencies to solve such problems successfully.

…going beyond the mastery of mathematical techniques conventionally taught at school.

This approach to mathematics contrasts with a traditional understanding of school mathematics which is often narrower. In schools, mathematical content is often taught and assessed in ways that are removed from authentic contexts – *e.g.,* students are taught the techniques of arithmetic, then given an arithmetic computation to complete; they are shown how to solve particular types of equations, then given further similar equations to solve; they are taught about geometric properties and relationships, then given a theorem to prove. Having learned the relevant concepts, skills and techniques, students are typically given contrived mathematical problems that call for the application of that knowledge. The mathematics required is usually obvious. Students have either mastered the techniques needed, or they have not. The usefulness of mathematics in the real world may be given little attention.

Assessment of such functional use of mathematics can influence how it is taught.

Outside school, real-life problems and situations for which mathematical knowledge may be useful often do not present themselves in such familiar forms. The individual must translate the situation or problem into a form that exposes the relevance and usefulness of mathematics. If students are unpractised at such a process, the potential power of mathematics to help deal with the situations and problems of their life may not be fully realised. The PISA approach to assessing mathematics was therefore designed to place the real-life use of mathematical knowledge and skills closer to the centre of a concept of mathematics learning. The intention is to encourage an approach to teaching and learning mathematics that gives strong emphasis to the processes associated with confronting problems in real-world contexts, making these problems amenable to mathematical treatment, using the relevant mathematical knowledge to solve problems, and evaluating the solution in the original problem context. If students can learn to do these things, they will be better equipped to make use of their mathematical knowledge and skills throughout life. They will be mathematically literate.

How mathematics is measured

PISA measures mathematics performance in three dimensions: mathematical content, the processes involved and the situations in which problems are posed.

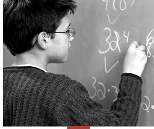
Students' mathematics knowledge and skills were assessed according to three dimensions relating to: the mathematical content to which different problems and questions relate; the processes that need to be activated in order to connect observed phenomena with mathematics and then to solve the respective problems; and the situations and contexts that are used as sources of stimulus materials and in which problems are posed.

Content

Tasks are divided into four areas of mathematical content.

PISA draws its mathematical content from broad content areas (OECD, 2003e). Taking account of the research literature on this subject, and following an in-depth consensus building process among OECD countries on what would be an appropriate basis to compare mathematics performance internationally, the assessment was established around four content areas:

- *Space and shape* relates to spatial and geometric phenomena and relationships, often drawing on the curricular discipline of geometry. It requires looking



for similarities and differences when analysing the components of shapes and recognising shapes in different representations and different dimensions, as well as understanding the properties of objects and their relative positions.

- *Change and relationships* involves mathematical manifestations of change as well as functional relationships and dependency among variables. This content area relates most closely to algebra. Mathematical relationships are often expressed as equations or inequalities, but relationships of a more general nature (*e.g.,* equivalence, divisibility and inclusion, to mention but a few) are relevant as well. Relationships are given a variety of different representations, including symbolic, algebraic, graphic, tabular and geometric representations. Since different representations may serve different purposes and have different properties, translation between representations is often of key importance in dealing with situations and tasks.
- *Quantity* involves numeric phenomena as well as quantitative relationships and patterns. It relates to the understanding of relative size, the recognition of numerical patterns, and the use of numbers to represent quantities and quantifiable attributes of real-world objects (counts and measures). Furthermore, quantity deals with the processing and understanding of numbers that are represented in various ways. An important aspect of dealing with quantity is quantitative reasoning, which involves number sense, representing numbers, understanding the meaning of operations, mental arithmetic and estimating. The most common curricular branch of mathematics with which quantitative reasoning is associated is arithmetic.
- *Uncertainty* involves probabilistic and statistical phenomena and relationships, that become increasingly relevant in the information society. These phenomena are the subject of mathematical study in statistics and probability.

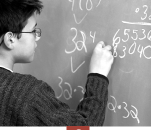
These relate to strands of the school curriculum…

Together, the four content areas cover the range of mathematics 15-year-olds need as a foundation for life and for further extending their horizon in mathematics. The concepts can be related to traditional content strands such as arithmetic, algebra or geometry and their detailed sub-topics that reflect historically well-established branches of mathematical thinking and that facilitate the development of a structured teaching syllabus.

…so performance reported separately on each content area can be related to countries' curricular choices.

The PISA mathematics assessment sets out to compare levels of student performance in these four content areas, with each area forming the basis of a scale reported later in this chapter. By reporting separately on student performance in each of four areas of mathematics, PISA recognises that different school systems choose to give different emphases in constructing their national curricula. Reporting in this way allows different school systems to situate their national priorities in relation to the choices made by other countries. It also allows different school systems to assess to what extent the level and growth of mathematical knowledge occur uniformly across these conceptually distinguishable assessment areas.

The first panel of Table A6.1 shows the breakdown by mathematical content area of the 85 test items used in the PISA 2003 assessment (Annex A6).



Process

To solve real-world problems, students must first transform them into a mathematical form, then perform mathematical operations, retranslate the result into the original problem and communicate the solution.

The PISA mathematics assessment requires students to confront mathematical problems that are based in some real-world context, where the students are required to identify features of the problem situation that might be amenable to mathematical investigation, and to activate the relevant mathematical competencies to solve the problem. In order to do so they need to engage in a multi-step process of "mathematisation": beginning with a problem situated in reality, students must organise it according to mathematical concepts. They must identify the relevant mathematical concepts, and then progressively trim away the reality in order to transform the problem into one that is amenable to direct mathematical solution, by making simplifying assumptions, by generalising and formalising information, by imposing useful ways of representing aspects of the problem, by understanding the relationships between the language of the problem and symbolic and formal language needed to understand it mathematically, by finding regularities and patterns and linking it with known problems or other familiar mathematical formulations and by identifying or imposing a suitable mathematical model.

This requires a number of different skills, which can be grouped in three categories…

Once the problem has been turned into a familiar or directly amenable mathematical form, the student's armoury of specific mathematical knowledge, concepts and skills can then be applied to solve it. This might involve a simple calculation, or using symbolic, formal and technical language and operations, switching between representations, using logical mathematical arguments, and generalising. The final steps in the mathematisation process involve some form of translation of the mathematical result into a solution that works for the original problem context, a reality check of the completeness and applicability of the solution, a reflection on the outcomes and communication of the results, which may involve explanation and justification or proof.

Various competencies are required for such mathematisation to be employed. These include: *thinking and reasoning; argumentation; communication; modelling; problem posing and solving; representation; and using symbolic, formal and technical language and operations*. While it is generally true that these competencies operate together, and there is some overlap in their definitions, PISA mathematics tasks were often constructed to call particularly on one or more of these competencies. The cognitive activities that the above mentioned competencies encompass were organised in PISA within three *competency clusters* that are labelled: the *reproduction cluster*, the *connections cluster*, and the *reflection cluster*. These groupings have been found to provide a convenient basis for discussing the way in which different competencies are invoked in response to the different kinds and levels of cognitive demands imposed by different mathematical problems.

…those involving familiar mathematical processes and computations…

- The *reproduction cluster* is called into play in those items that are relatively familiar, and that essentially require the reproduction of practised knowledge, such as knowledge of facts and of common problem representations, recognition of equivalents, recollection of familiar mathematical objects and properties,

performance of routine procedures, application of standard algorithms and technical skills, manipulation of expressions containing symbols and formulae in a familiar and standard form, and carrying out straight-forward computations.

- The *connections cluster* builds on reproduction to solve problems that are not simply routine, but that still involve somewhat familiar settings or extend and develop beyond the familiar to a relatively minor degree. Problems typically involve greater interpretation demands, and require making links between different representations of the situation, or linking different aspects of the problem situation in order to develop a solution.

…those involving a degree of interpretation and linkages…

- The *reflection cluster* builds further on the connections cluster. These competencies are required in tasks that demand some insight and reflection on the part of the student, as well as creativity in identifying relevant mathematical concepts or in linking relevant knowledge to create solutions. The problems addressed using the competencies in this cluster involve more elements than others, and additional demands typically arise for students to generalise and to explain or justify their results.

…and those involving deeper insights and reflection.

The second panel in Table A6.1 shows the breakdown by competency cluster of the 85 test items used in the PISA 2003 assessment (Annex A6). A more detailed description of these competency clusters and the ways in which the individual competencies operate in each of these clusters is described in *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills* (OECD, 2003e).

Situation

As in PISA 2000, students were shown various pieces of written material, and for each were asked a series of questions. The stimulus material represented a situation that students could conceivably confront, and for which activation of their mathematical knowledge, understanding or skill might be required or might be helpful in order to analyse or deal with the situation. There were of four sorts of situations: personal, educational or occupational, public and scientific.

PISA mathematics tasks are set in a range of contexts, relating to…

- *Personal situations* directly relate to students' personal day-to-day activities. These have at their core the way in which a mathematical problem immediately affects the individual and the way the individual perceives the context of the problem. Such situations tend to require a high degree of interpretation before the problem can be solved.

…day-to-day activities…

- *Educational or occupational* situations appear in a student's life at school, or in a work setting. These have at their core the way in which the school or work setting might require a student or employee to confront some particular problem that requires a mathematical solution.

…school and work situations…

- *Public situations relating to the local and broader community* require students to observe some aspect of their broader surroundings. These are generally situations located in the community that have at their core the way in which

…the wider community…

students understand relationships among elements of their surroundings. They require the students to activate their mathematical understanding, knowledge and skills to evaluate aspects of an external situation that might have some relevant consequences for public life.

…and scientific or explicitly mathematical problems.

- *Scientific situations* are more abstract and might involve understanding a technological process, theoretical situation or explicitly mathematical problem. The PISA mathematics framework includes in this category relatively abstract mathematical situations with which students are frequently confronted in a mathematics classroom, consisting entirely of explicit mathematical elements and where no attempt is made to place the problem in some broader context. These are sometimes referred to as "intra-mathematical" contexts.

These situations differ in terms of how directly the problem affects students' lives…

These four situation types vary in two important respects. The first is in terms of the distance between the student and the situation – the degree of immediacy and directness of the problem's impact on the student. Personal situations are closest to students, being characterised by the direct perceptions involved. Educational and occupational situations typically involve some implications for the individual through their daily activities. Situations relating to the local and broader community typically involve a slightly more removed observation of external events in the community. Finally, scientific situations tend to be the most abstract and therefore involve the greatest separation between the student and the situation. The PISA assessment assumes that students need to be able to handle a range of situations, both close to and distant from their immediate lives.

…and also in the extent to which the mathematical aspects are explicit.

There are also differences in the extent to which the mathematical nature of a situation is apparent. A few of the tasks refer only to mathematical objects, symbols or structures, and make no reference to matters outside the mathematical world. However, PISA also encompasses problems that students are likely to encounter in their lives in which the mathematical elements are not stated explicitly. The assessment thus tests the extent to which students can identify mathematical features of a problem when it is presented in a non-mathematical context and the extent to which they can activate their mathematical knowledge to explore and solve the problem and to make sense of the solution in the context or situation in which the problem arose.

The third panel of Table A6.1 shows the breakdown by situation type of the 85 test items used in the PISA 2003 assessment (Annex A6).

A more detailed description of the conceptual underpinning of the PISA 2003 assessment as well as the characteristics of the test itself can be found in *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills* (OECD, 2003e).

How the PISA tests were constructed

Experts developed tasks designed to cover the PISA framework…

Assessment items were constructed to cover the different dimensions of the PISA assessment framework described above. During the process of item development, experts from participating countries undertook a qualitative

analysis of each item, and developed descriptions of aspects of the cognitive demands of each item. This analysis included judgements about the aspects of the PISA mathematics framework that were relevant to the item. A short description was then developed that captured the most important demands placed on students by each particular item, particularly the individual competencies that were called into play (*PISA 2003 Technical Report*, OECD, forthcoming).

…some requiring open answers that were scored by expert markers in a process involving intra-country and inter-country reliability checks…

The items had a variety of formats. In many cases, students were required to construct a response in their own words to questions based on the text given. Sometimes they had to write down their calculations in order to demonstrate some of the methods and thought processes they used in producing an answer. Other questions required students to write an explanation of their results, which again exposed aspects of the methods and thought processes they had employed to answer the question. These open-constructed response items could not easily be machine-scored; rather they required the professional judgement of trained markers to assign the observed responses to defined response categories. To ensure that the marking process yielded reliable and cross-nationally comparable results, detailed guidelines and training contributed to a marking process that was accurate and consistent across countries. In order to examine the consistency of this marking process in more detail within each country and to assess the consistency in the work of the markers, a subsample of items in each country was rated independently by four markers. The PISA Consortium then assessed the reliability of these markings. Finally, to verify that the marking process was carried out in equivalent ways across countries, an inter-country reliability study was carried out on a subset of items. In this process, independent marking of the original booklets was undertaken by trained multilingual staff and compared to the ratings by the national markers in the various countries. The results show that very consistent marking was achieved across countries (Annex A7; *PISA 2003 Technical Report*, OECD, forthcoming).

…but computers could mark tasks with a more limited set of possible responses…

For other items requiring students to construct a response, evaluation of their answers was restricted to the response itself rather than an explanation of how it was derived. For many of these closed constructed-response items, the answer given was in numeric or other fixed form, and could be evaluated against precisely defined criteria. Such responses generally did not require expert markers, but could be analysed by computer.

…including those where students had to choose from stated options.

Items that required students to select one or more responses from a number of given possible answers were also used. This format category includes both standard multiple-choice items, for which students were required to select one correct response from a number of given response options; and complex multiple-choice items, for which students were required to select a response from given optional responses to each of a number of propositions or questions. Responses to these items could be marked automatically.

Table A6.1 shows the breakdown by item format type of the 85 test items used in the PISA 2003 assessment (Annex A6).

Students were given credit for each item that they answered with an acceptable response. In the development of the assessment, extensive field trials were carried out in all participating countries in the year prior to the assessment to identify and anticipate the widest possible range of student responses. These were then assigned to distinct categories by the item developers to determine marks. In some cases, where there is clearly a correct answer, responses can be easily identified as being correct or not. In other cases a range of different responses might be regarded as being correct. In yet other cases, a range of different responses can be identified and among those some are clearly better than others. In such cases it is often possible to define several response categories that are ordered in their degree of correctness – one kind of response is clearly best, a second category is not quite as good but is better than a third category, and so on. In these cases partial credit could be given.

How the PISA tests were designed, analysed and scaled

Each student was given a subset from a broad pool of mathematics tasks…

In total, 85 mathematics items were used in PISA 2003. These tasks, and also those in reading, science and problem solving, were arranged into half-hour clusters. Each student was given a test booklet with four clusters of items – resulting in two hours of individual assessment time. These clusters were rotated in combinations that ensured that each mathematics item appeared in the same number of test booklets, and that each cluster appeared in each of the four possible positions in the booklets.

…and their performance was established on a scale…

Such a design makes it possible to construct a scale of mathematical performance, to associate each assessment item with a point score on this scale according to its difficulty and to assign each student a point score on the same scale representing his or her estimated ability. This is possible using techniques of modern item response modelling (a description of the model can be found in the *PISA 2003 Technical Report*, OECD, forthcoming).

The relative ability of students taking a particular test can be estimated by considering the proportion of test items they answer correctly. The relative difficulty of items in a test can be estimated by considering the proportion of test takers getting each item correct. The mathematical model employed to analyse the PISA data was implemented through iterative procedures that simultaneously estimate the probability that a particular person will respond correctly to a given set of test items, and the probability that a particular item will be answered correctly by a given set of students. The result of these procedures is a set of estimates that allows the creation of a continuous scale representing mathematical literacy. On this continuum it is possible to estimate the location of individual students, thereby seeing what degree of mathematical literacy they demonstrate, and it is possible to estimate the location of individual test items, thereby seeing what degree of mathematical literacy each item embodies.[4]

Once the difficulty of individual items was given a rating on the scale, student performance could be described by giving each student a score according to

the hardest task that they could be predicted to perform. This does not mean that students will *always* be able to perform items at or below the difficulty level associated with their own position on the scale, and *never* be able to do harder items. Rather, the ratings are based on probability. As illustrated in Figure 2.1, students have a relatively high probability[5] of being able to complete items below their own rating (with the probability rising for items further down the scale), but are relatively unlikely to be able to complete those items further up.

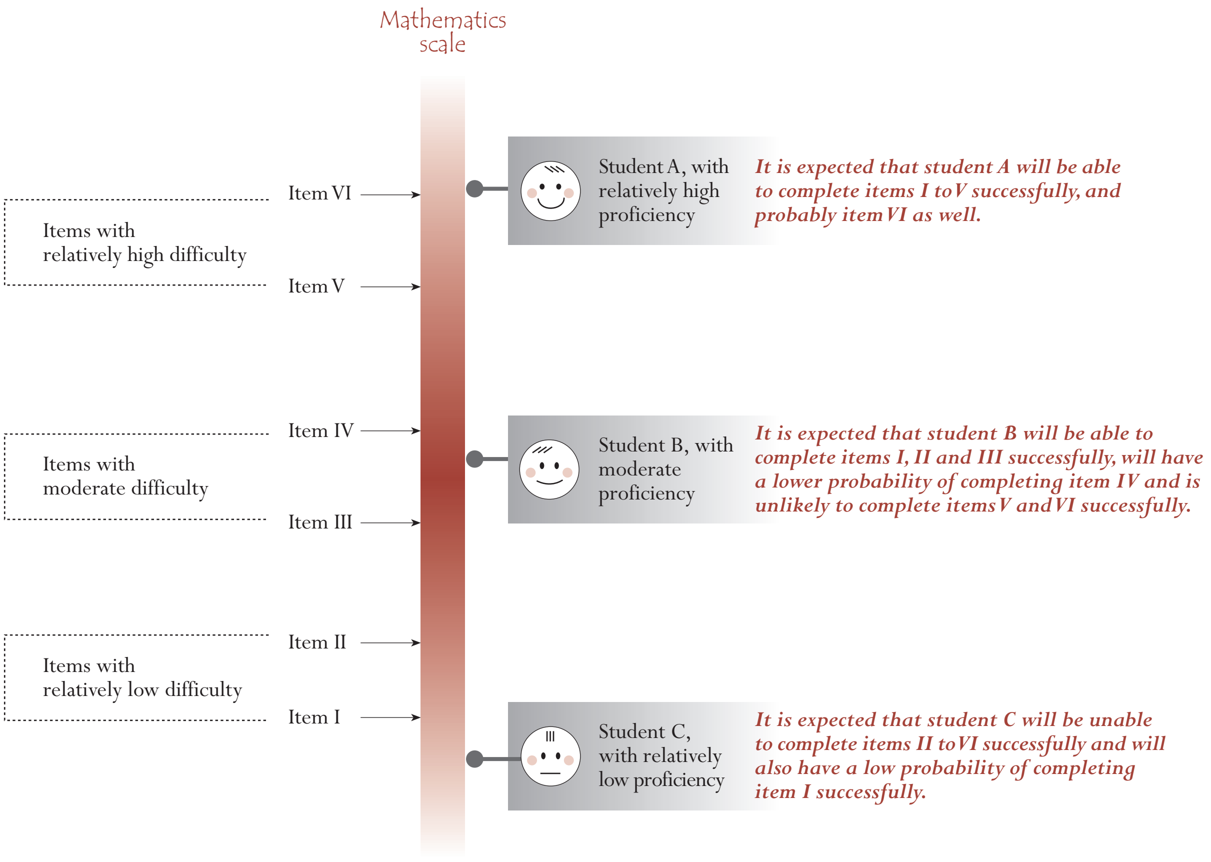
…with a score of 500 representing average OECD performance.

To facilitate the interpretation of the scores assigned to students, the scale was constructed to have an average score among OECD countries of 500 points, with about two-thirds of students across OECD countries scoring between 400 and 600 points.[6]

Students were grouped in six levels of proficiency, plus a group below Level 1…

In a manner similar to the reporting of the PISA 2000 reading assessment, which presented results in proficiency levels, student scores in mathematics in 2003 were grouped into six proficiency levels. The six proficiency levels represented groups of tasks of ascending difficulty, with Level 6 as the highest and Level 1 as the lowest. The grouping into proficiency levels was undertaken on the

Figure 2.1 ■ **The relationship between items and students on a proficiency scale**



basis of substantive considerations relating to the nature of the underlying competencies. Students with below 358 score points on any of the mathematics scales were classified as below Level 1. Such students, representing 11 per cent of students on average across OECD countries, were not necessarily incapable of performing any mathematical operation. However, they were unable to utilise mathematical skills in the situations required by the easiest PISA tasks.

...with each proficiency level relating to a specific set of mathematical competencies.

Proficiency at each of these levels can be understood in relation to descriptions of the kind of mathematical competency that a student needs to attain them. These are summarised in Figure 2.2. In fact, these descriptions represent a synthesis of the proficiency descriptions for each of the content areas of mathematics, which are given later in this chapter when discussing results in each content area. The progression through these levels, in terms of the ways in which the individual mathematical processes change as levels increase is shown in Annex A2.

The creation of the six proficiency levels leads to a situation where students with a range of scores on a continuous scale are grouped together into each single band. PISA applies an easy-to-understand criterion to assigning students to levels: each student is assigned to the highest level for which they would be expected to answer correctly the majority of assessment items. Thus, for example, in a test composed of items spread uniformly across Level 3 (with difficulty ratings of 483 to 544 score points), all students assigned to that level would expect to get at least 50 per cent of the items correct. Someone at the bottom of the level (scoring 483 points) would be expected to get close to 50 per cent of the items correct; someone in the middle or near the top of the level would get a higher percentage of items correct. For this to be true, a student scoring 483 needs to have a 50 per cent chance of completing an item in the middle of Level 3 (rated 513 score points) and thus have a greater than 50 per cent chance of getting right an item rated at their score, 483 points. This latter probability needs to be 62 per cent to fulfil these conditions.

How results are reported

PISA 2003 mathematics results are reported on four scales relating to the content areas described above. Performance is also reported on an overall mathematics scale.

The mathematics tasks can be mapped according to difficulty...

Figure 2.3 shows a map with a sample of items from the PISA 2003 assessment, with the items shown in detail in Figures 2.4a-c, Figures 2.7a-b, Figures 2.10a-b and Figures 2.13a-c. For each of the four content areas, the selected items and item scores (*i.e.,* full or partial credit) have been ordered according to their difficulty, with the most difficult of these scores at the top, and the least difficult at the bottom.

...with the easiest tasks tending to require mainly reproduction skills and the hardest ones reflection.

The characteristics of the items shown in the map provide the basis for a substantive interpretation of performance at different levels on the scale. Patterns emerge that make it possible to describe aspects of mathematics that are consistently associated with various locations along the literacy continuum shown by the map. For example, among the small sample of items in Figure 2.3,

Figure 2.2 ■ **Summary descriptions for the six levels of proficiency in mathematics**

Level	WHAT STUDENTS CAN TYPICALLY DO
6	At Level 6, students can conceptualise, generalise, and utilise information based on their investigations and modelling of complex problem situations. They can link different information sources and representations and flexibly translate among them. Students at this level are capable of advanced mathematical thinking and reasoning. These students can apply this insight and understanding, along with a mastery of symbolic and formal mathematical operations and relationships, to develop new approaches and strategies for attacking novel situations. Students at this level can formulate and precisely communicate their actions and reflections regarding their findings, interpretations, arguments, and the appropriateness of these to the original situations.
5	At Level 5, students can develop and work with models for complex situations, identifying constraints and specifying assumptions. They can select, compare, and evaluate appropriate problem-solving strategies for dealing with complex problems related to these models. Students at this level can work strategically using broad, well-developed thinking and reasoning skills, appropriately linked representations, symbolic and formal characterisations, and insight pertaining to these situations. They can reflect on their actions and can formulate and communicate their interpretations and reasoning.
4	At Level 4, students can work effectively with explicit models for complex concrete situations that may involve constraints or call for making assumptions. They can select and integrate different representations, including symbolic ones, linking them directly to aspects of real-world situations. Students at this level can utilise well-developed skills and reason flexibly, with some insight, in these contexts. They can construct and communicate explanations and arguments based on their interpretations, arguments and actions.
3	At Level 3, students can execute clearly described procedures, including those that require sequential decisions. They can select and apply simple problem-solving strategies. Students at this level can interpret and use representations based on different information sources and reason directly from them. They can develop short communications reporting their interpretations, results and reasoning.
2	At Level 2, students can interpret and recognise situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formulae, procedures or conventions. They are capable of direct reasoning and making literal interpretations of the results.
1	At Level 1, students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are obvious and follow immediately from the given stimuli.

Figure 2.3 ■ A map of selected mathematics items

Level	Space and shape Figures 2.4a-c	Change and relationships Figures 2.7a-b	Quantity Figures 2.10a-b	Uncertainty Figures 2.13a-c
6	**CARPENTER** *Question 1 (687)*	**WALKING** *Question 5 – Score 3 (723)*		**ROBBERIES** *Question 15 – Score 2 (694)*
668.7				
5		**WALKING** *Question 5 – Score 2 (666)* **WALKING** *Question 4 (611)*		**TEST SCORES** *Question 6 (620)*
606.6				
4		**WALKING** *Question 5 – Score 1 (605)* **GROWING UP** *Question 8 (574)*	**EXCHANGE RATE** *Question 11 (586)* **SKATEBOARD** *Question 13 (570)* **SKATEBOARD** *Question 14 (554)*	**ROBBERIES** *Question 15 – Score 1 (577)* **EXPORTS** *Question 18 (565)*
544.4				
3	**NUMBER CUBES** *Question 3 (503)*	**GROWING UP** *Question 7 – Score 2 (525)*	**SKATEBOARD** *Question 12 – Score 2 (496)*	OECD average = 500
482.4				
2	**STAIRCASE** *Question 2 (421)*		**SKATEBOARD** *Question 12 – Score 1 (464)* **EXCHANGE RATE** *Question 10 (439)*	**EXPORTS** *Question 17 (427)*
420.4				
1		**GROWING UP** *Question 7 – Score 1 (420)*	**EXCHANGE RATE** *Question 9 (406)*	
358.3				
Below Level 1				

the easiest items are all from the *reproduction* competency cluster. This reflects the pattern observed with the full set of items. It is also seen from the full set of PISA items that those items characterised as belonging to the *reflection* cluster tend to be the most difficult. Items in the *connections* cluster tend to be of intermediate difficulty, though they span a large part of the proficiency spectrum that is analysed through the PISA assessment. The individual competencies defined in the mathematics framework operate quite differently at different levels of performance, as predicted by the assessment framework.

The easiest tasks require straightforward mathematical operations in familiar contexts…

Near the bottom of the scale, items set in simple and relatively familiar contexts require only the most limited interpretation of the situation, as well as direct application of well-known mathematical knowledge in familiar situations. Typical activities are reading a value directly from a graph or table, performing a very simple and straightforward arithmetic calculation, ordering a small set of numbers correctly, counting familiar objects, using a simple currency exchange rate, identifying and listing simple combinatorial outcomes. For example, Question 9 from the unit *Exchange Rate* (Figure 2.10a) presents students with a simple rate for exchanging Singapore dollars (SGD) into South African rand (ZAR), namely 1 SGD = 4.2 ZAR. The question requires students to apply the rate to convert 3000 SGD into ZAR. The rate is presented in the form of a familiar equation, and the mathematical step required is direct and reasonably obvious. In examples 9.1 and 9.2 from the unit *Building Blocks* (OECD, 2003e), students were presented with diagrams of familiar three-dimensional shapes composed of small cubes, and asked to count (or calculate) the number of the small cubes used to make up the larger shapes.

…and tasks of medium difficulty require more transformation into mathematical form…

Around the middle of the scale, items require substantially more interpretation, frequently of situations that are relatively unfamiliar or unpractised. They often demand the use of different representations of the situation, including more formal mathematical representations, and the thoughtful linking of those different representations in order to promote understanding and facilitate analysis. They often involve a chain of reasoning or a sequence of calculation steps, and can require students to express reasoning through a simple explanation. Typical activities include interpreting a set of related graphs; interpreting text, relating this to information in a table or graph, extracting the relevant information and performing some calculations; using scale conversions to calculate distances on a map; and using spatial reasoning and geometric knowledge to perform distance, speed and time calculations. For example, the unit *Growing Up* (Figure 2.7b) presents students with a graph of the average height of young males and young females from the ages of ten to 20 years. Question 7 from *Growing Up* asks students to identify the period in their life when females are on average taller than males of the same age. Students have to interpret the graph to understand exactly what is being displayed. They also have to relate the graphs for males and females to each other and determine how the specified period is shown then accurately read the relevant values from the horizontal scale. Question 8 from the unit *Growing Up* invites students to give a written explanation as to how the

graph shows a slowdown in the growth rate for girls after a particular age. To answer this question successfully, students must understand how the growth rate is displayed in such a graph, identify what is changing at the specified point in the graph in comparison to an earlier period and clearly articulate their explanation in words.

...while difficult tasks are more complex and require greater interpretation of unfamiliar problems.

Towards the top of the scale, items are displayed that typically involve a number of different elements, and require even higher levels of interpretation. Situations are typically unfamiliar, hence requiring some degree of thoughtful reflection and creativity. Questions usually demand some form of argument, often in the form of an explanation. Typical activities involved include: interpreting complex and unfamiliar data; imposing a mathematical construction on a complex real-world situation; and using mathematical modelling processes. At this part of the scale, items tend to have several elements that need to be linked by students, and their successful negotiation typically requires a strategic approach to several interrelated steps. For example, Question 15 from the unit *Robberies* (Figure 2.13a) presents students with a truncated bar graph showing the number of robberies per year in two specified years. A television reporter's statement interpreting the graph is given. Students are asked to consider whether or not the reporter's statement is a reasonable interpretation of the graph, and to give an explanation as to why. The graph itself is somewhat unusual, and requires some interpretation. The reporter's statement must be interpreted in relation to the graph. Then, some mathematical understanding and reasoning must be applied to determine a suitable meaning of the phrase "reasonable interpretation" in this context. Finally, the conclusion must be articulated clearly in a written explanation. Fifteen-year-old students typically find such a sequence of thought and action quite challenging.

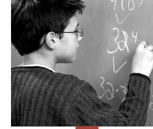
Another example presented in the PISA assessment framework, example 3.2 in the unit *Heartbeat* (OECD, 2003e), presents students with mathematical formulations of the relationship between a person's recommended maximum heart rate and their age, in the context of physical exercise. The question invites students to modify the formulation appropriately under a specified condition. They have to interpret the situation, the mathematical formulations, the changed condition, and construct a modified formulation that satisfies the specified condition. This complex set of linked tasks also proved to be very demanding for 15-year-olds.

Thus, difficulty rises with the amount of interpretation, representation, complex processing and argumentation required of students.

Based on the patterns observed when the full item set is investigated in this way, it is possible to characterise growth along the PISA mathematics scale by referring to the ways in which mathematical competencies are associated with items located at different points along the scale.

The ascending difficulty of mathematics items is associated with:

- The kind and degree of interpretation and reflection needed, including the nature of demands arising from the problem context; the extent to which the mathematical demands of the problem are apparent or to which students must



impose their own mathematical construction on the problem; and the extent to which insight, complex reasoning and generalisation are required.

- The kind of representation skills that are necessary, ranging from problems where only one mode of representation is used to problems where students have to switch between different modes of representation or to find appropriate modes of representation themselves.
- The kind and level of mathematical complexity required, ranging from single-step problems requiring students to reproduce basic mathematical facts and perform simple computation processes through to multi-step problems involving more advanced mathematical knowledge, complex decision-making, information processing, problem-solving and modelling skills.
- The kind and degree of mathematical argumentation that is required, ranging from problems where no argumentation is necessary at all, through to problems where students may apply well-known arguments, to problems where students have to create mathematical arguments or to understand other people's argumentation or judge the correctness of given arguments or proofs.

WHAT STUDENTS CAN DO IN FOUR AREAS OF MATHEMATICS

By looking at how students performed on the four scales, alongside examples of the tasks associated with those content areas of mathematics, it is possible to provide a profile of what PISA shows about students' mathematical abilities. For two of these areas – *change and relationships* and *space and shape*, it is also possible to compare mathematical performance in 2003 with that measured in PISA 2000.

Student performance can be summarised on four scales, relating to space and shape, change and relationships, quantity, and uncertainty phenomena.

Student performance on the mathematics/space and shape scale

A quarter of the mathematical tasks given to students in PISA are related to spatial and geometric phenomena and relationships. Figures 2.4a-c show three sample tasks from this category: one at Level 2, one at Level 3 and one at Level 6.

The knowledge and skills required to reach each level are summarised in Figure 2.5. In PISA 2003, only a small proportion of 15-year-olds – 5 per cent overall in the combined OECD area[7] – can perform the highly complex tasks required to reach Level 6. However, more than 15 per cent of the students in Korea and the PISA partner country Hong Kong-China, and more than 10 per cent of the students in Belgium, the Czech Republic, Japan and Switzerland as well as the partner country Liechtenstein (Figure 2.6a) perform at Level 6. In contrast, in Greece, Mexico and Portugal, as well as in the partner countries Brazil, Indonesia, Serbia,[8] Thailand, Tunisia and Uruguay, less than 1 per cent reach Level 6 (Table 2.1a).

In most countries under 10 per cent of students can perform the hardest space and shape tasks…

A quarter or more of students fail to reach Level 2 in Greece, Hungary, Ireland, Italy, Luxembourg, Mexico, Norway, Poland, Portugal, Spain, Turkey and the United States as well as in the partner countries Brazil, Indonesia, Latvia, the Russian Federation, Serbia, Thailand, Tunisia and Uruguay.

…but in 12 OECD countries at least 25 per cent can only perform very simple tasks.

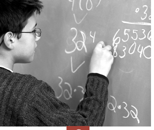
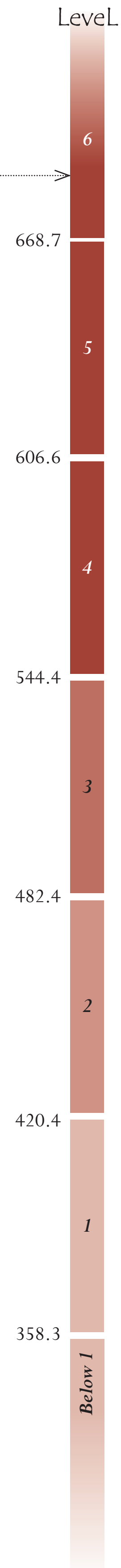
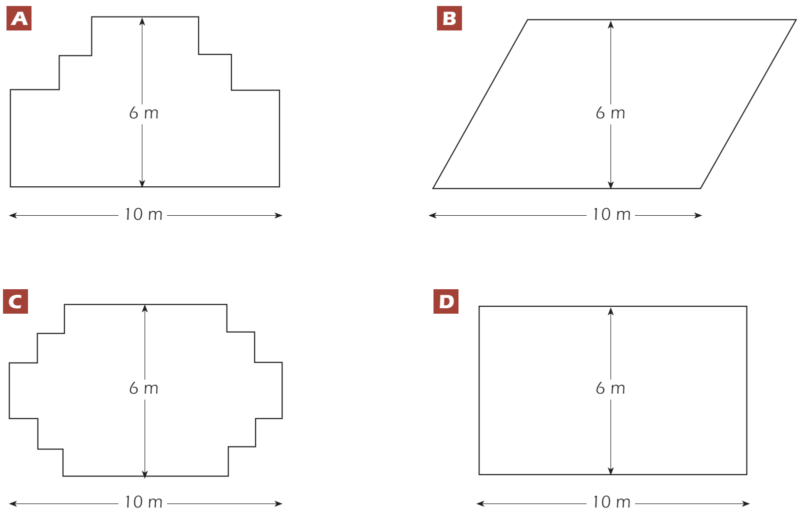


Figure 2.4a ■ **A sample of mathematics items used in PISA for the space and shape scale: Unit CARPENTER**



CARPENTER

A carpenter has 32 metres of timber and wants to make a border around a garden bed. He is considering the following designs for the garden bed.



QUESTION 1

Circle either "Yes" or "No" for each design to indicate whether the garden bed can be made with 32 metres of timber.

Garden bed design	Using this design, can the garden bed be made with 32 metres of timber?
Design A	*Yes / No*
Design B	*Yes / No*
Design C	*Yes / No*
Design D	*Yes / No*

Score 1 (687)

Answers which indicate Yes, No, Yes, Yes, in that order.

This complex multiple-choice item is situated in an educational context, since it is the kind of quasi-realistic problem that would typically be seen in a mathematics class, rather than being a genuine problem likely to be met in an occupational setting. While not regarded as typical, a small number of such problems have been included in the PISA assessment. However, the competencies needed for this problem are certainly relevant and part of mathematical literacy. This item illustrates Level 6 with a difficulty of 687 score points. The item belongs to the space and shape content area, and it fits the connections competency cluster – as the problem is non-routine. The students need the competence to recognise that for the purpose of solving the question the two-dimensional shapes A, C and D have the same perimeter, therefore they need to decode the visual information and see similarities and differences. The students need to see whether or not a certain border-shape can be made with 32 metres of timber. In three cases this is rather evident because of the rectangular shapes. But the fourth is a parallelogram, requiring more than 32 metres. This use of geometrical insight and argumentation skills and some technical geometrical knowledge makes this item illustrate the Level 6.

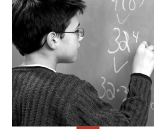
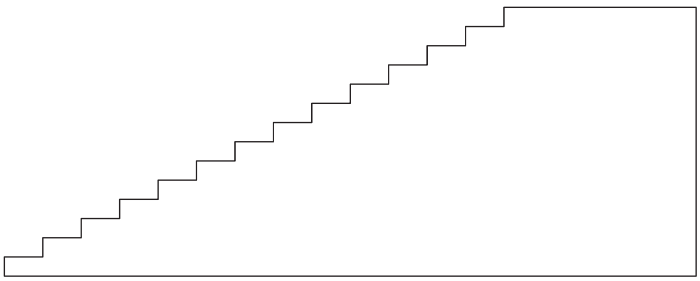


Figure 2.4b ■ **A sample of mathematics items used in PISA for the space and shape scale: Unit STAIRCASE**

STAIRCASE

The diagram below illustrates a staircase with 14 steps and a total height of 252 cm:



QUESTION 2

What is the height of each of the 14 steps?
Height: ....cm.

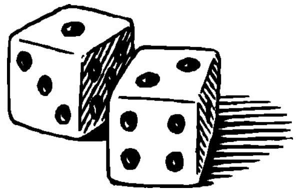
Score 1 (421)
Answers which indicate 18 cm.

This short open-constructed response item is situated in a daily life context for carpenters and therefore is classified as having an occupational context. It has a difficulty of 421 score points. One does not need to be a carpenter to understand the relevant information; it is clear that an informed citizen should be able to interpret and solve a problem like this that uses two different representation modes: language, including numbers, and a graphical representation. But the illustration serves a simple and non-essential function: students know what stairs look like. This item is noteworthy because it has redundant information (the depth is 400 cm) that is sometimes considered by students as confusing, but such redundancy is common in real-world problem solving. The context of the stairs places the item in the space and shape content area, but the actual procedure to carry out is a simple division. As this is a basic operation with numbers (divide 252 by 14) the item belongs to the reproduction competency cluster. The problem-solving competency involved here solving problems by invoking and using standard approaches and procedures in one way only. All the required information, and even more than required, is presented in a recognisable situation, the students can extract the relevant information from a single source, and, in essence the item makes use of a single representational mode. Combined with the application of a basic algorithm makes this item fit, although barely, at Level 2.

LeveL
6
668.7
5
606.6
4
544.4
3
482.4
2
420.4
1
358.3
Below 1

Figure 2.4c ■ **A sample of mathematics items used in PISA for the space and shape scale: Unit NUMBER CUBES**

NUMBER CUBES

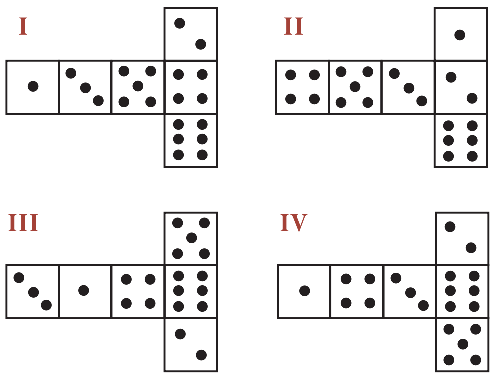


On the right, there is a picture of two dice.
Dice are special number cubes for which the following rule applies:
"The total number of dots on two opposite faces is always seven."

QUESTION 3

You can make a simple number cube by cutting, folding and gluing cardboard. This can be done in many ways. In the figure below you can see four cuttings that can be used to make cubes, with dots on the sides.

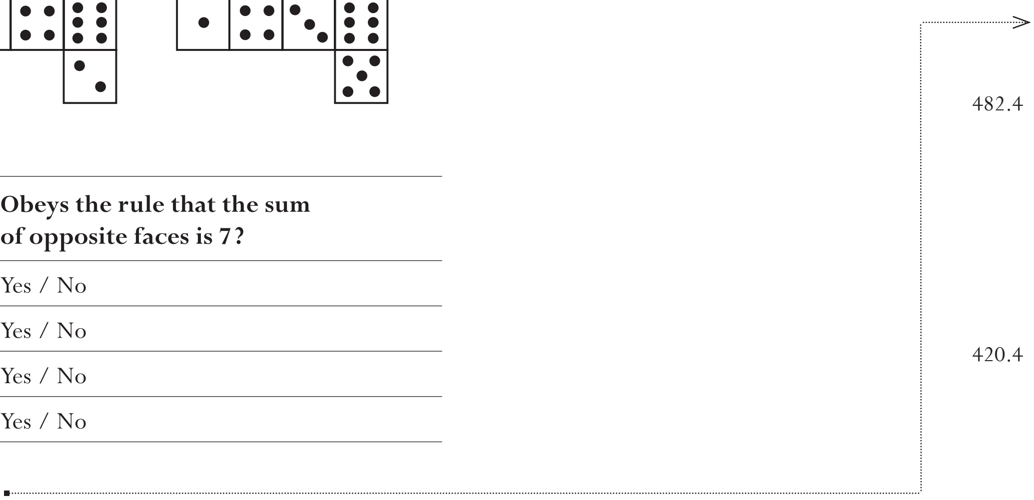
Which of the following shapes can be folded together to form a cube that obeys the rule that the sum of opposite faces is 7? For each shape, circle either "Yes" or "No" in the table below.



Shape	Obeys the rule that the sum of opposite faces is 7?
I	Yes / No
II	Yes / No
III	Yes / No
IV	Yes / No

Score 1 (503)

Answers which indicate No, Yes, Yes, No, in that order.



This complex multiple-choice item is situated in a personal context. It has a difficulty of 503 score points. Many games that children encounter during their education, whether formal or informal, use number cubes. The problem does not assume any previous knowledge about this cube, but an understanding of the rule of its construction: two opposite sides have a total of seven dots. This construction rule emphasises a numerical aspect, but the problem posed requires some kind of spatial insight or mental visualisation technique. These competencies are an essential part of mathematical literacy as students live in three-dimensional space, and often are confronted with two-dimensional representations. Students need to mentally imagine how the four plans of number cubes, if reconstructed into a 3-D number cube, obey the numerical construction rule. Therefore the item belongs to the space and shape content area. The problem is not routine: it requires the encoding and spatial interpretation of two-dimensional objects, interpretation of the connected three-dimensional object, interpreting back-and-forth between model and reality, and checking certain basic quantitative relations. This leads to a classification in the connections competency cluster. The item requires spatial reasoning skills within a personal context with all the relevant information clearly presented in writing and with graphics. The item illustrates Level 3.

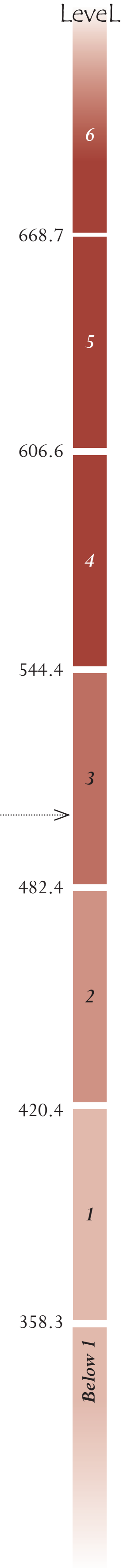
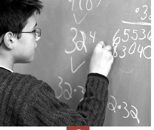


Figure 2.5 ■ **Summary descriptions of six levels of proficiency on the mathematics/space and shape scale**

Level	General competencies students should have at each level	Specific tasks students should be able to do
6	*5% of all students across the OECD area can perform tasks at Level 6 on the space and shape scale*	
	Solve complex problems involving multiple representations and often involving sequential calculation processes; identify and extract relevant information and link different but related information; use reasoning, significant insight and reflection; and generalise results and findings, communicate solutions and provide explanations and argumentation	– Interpret complex textual descriptions and relate these to other (often multiple) representations – Use reasoning involving proportions in non-familiar and complex situations – Show significant insight to conceptualise complex geometric situations or to interpret complex and unfamiliar representations – Identify and combine multiple pieces of information to solve problems – Devise a strategy to connect a geometrical context with known mathematical procedures and routines – Carry out a complex sequence of calculations, for example volume calculations or other routine procedures in an applied context, accurately and completely – Provide written explanations and arguments based on reflection, insight and generalisation of understanding
5	*15% of all students across the OECD area can perform tasks at least at Level 5 on the space and shape scale*	
	Solve problems that require appropriate assumptions to be made, or that involve working with assumptions provided; use well-developed spatial reasoning, argument and insight to identify relevant information and to interpret and link different representations; work strategically and carry out multiple and sequential processes	– Use spatial/geometrical reasoning, argument, reflection and insight into two- and three-dimensional objects, both familiar and unfamiliar – Make assumptions or work with assumptions to simplify and solve a geometrical problem in a real-world setting, *e.g.,* involving estimation of quantities in a real-world situation, and communicate explanations – Interpret multiple representations of geometric phenomena – Use geometric constructions – Conceptualise and devise multi-step strategies to solve geometrical problems – Use well-known geometrical algorithms but in unfamiliar situations, such as Pythagoras' theorem, and calculations involving perimeter, area and volume
4	*30% of all students across the OECD area can perform tasks at least at Level 4 on the space and shape scale*	
	Solve problems that involve visual and spatial reasoning and argumentation in unfamiliar contexts; link and integrate different representations; carry out sequential processes; apply well-developed skills in spatial visualisation and interpretation	– Interpret complex text to solve geometric problems – Interpret sequential instructions and follow a sequence of steps – Interpretation using spatial insight into non–standard geometric situations – Use a two–dimensional model to work with 3-D representations of unfamiliar geometric situation – Link and integrate two different visual representations of geometric situations – Develop and implement a strategy involving calculation in geometric situations – Reason and argue about numeric relationships in a geometric context – Perform simple calculations (*e.g.,* multiply multi-digit decimal number by an integer, apply numeric conversions using proportion and scale, calculate areas of familiar shapes)

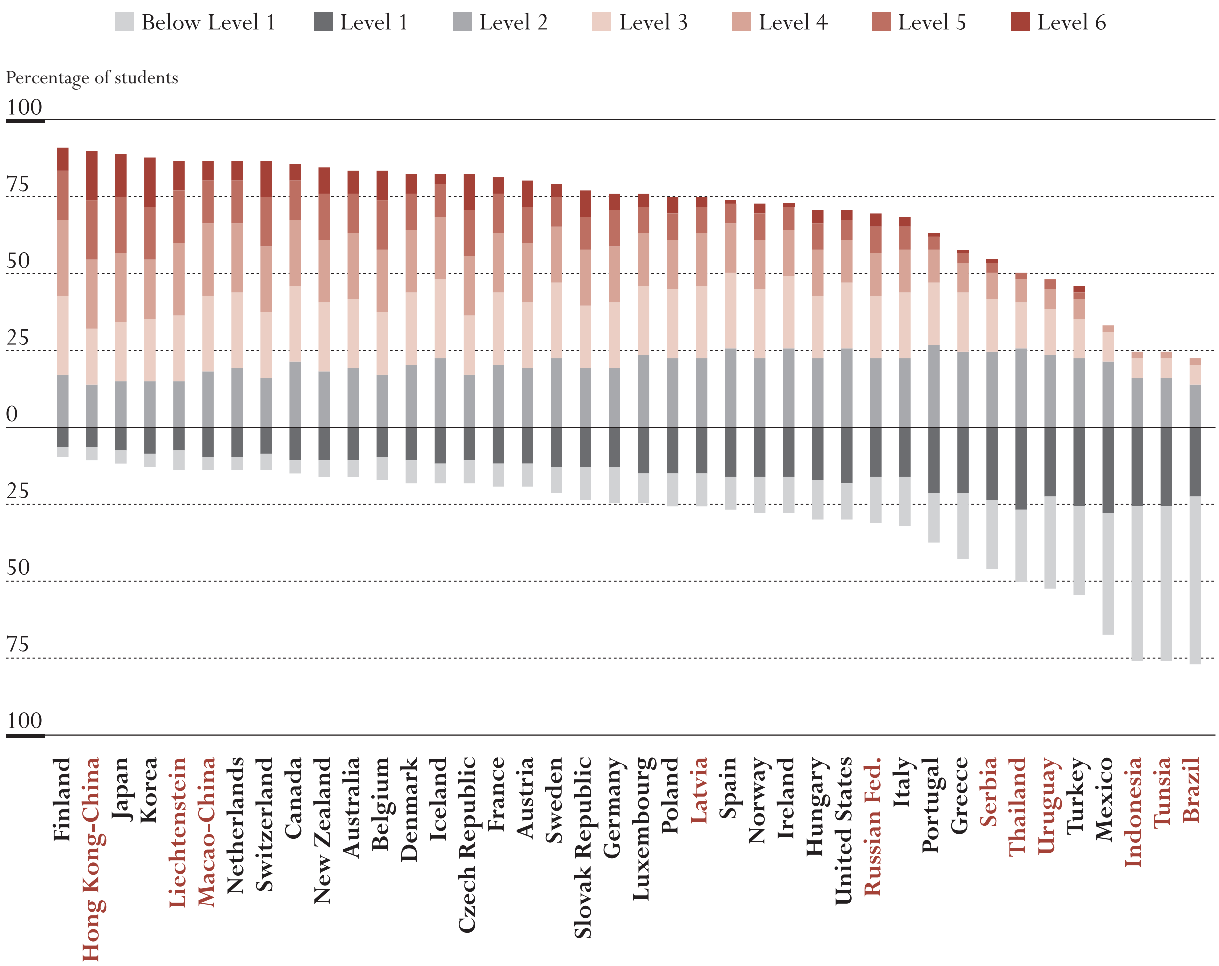


Level	General competencies students should have at each level	Specific tasks students should be able to do
3	*51% of all students across the OECD area can perform tasks at least at Level 3 on the space and shape scale*	
	Solve problems that involve elementary visual and spatial reasoning in familiar contexts; link different representations of familiar objects; use elementary problem solving skills (devising simple strategies); apply simple algorithms	– Interpret textual descriptions of unfamiliar geometric situations – Use basic problem–solving skills, such as devising a simple strategy – Use visual perception and elementary spatial reasoning skills in a familiar situation – Work with a given familiar mathematical model – Perform simple calculations such as scale conversions (using multiplication, basic proportional reasoning) – Apply routine algorithms to solve geometric problems (*e.g.,* calculate lengths within familiar shapes)
2	*71% of all students across the OECD area can perform tasks at least at Level 2 on the space and shape scale*	
	Solve problems involving a single mathematical representation where the mathematical content is direct and clearly presented; use basic mathematical thinking and conventions in familiar contexts	– Recognise simple geometric patterns – Use basic technical terms and definitions and apply basic geometric concepts (*e.g.,* symmetry) – Apply a mathematical interpretation of a common-language relational term (*e.g.,* "bigger") in a geometric context – Create and use a mental image of an object, both two- and three-dimensional – Understand a visual two-dimensional representation of a familiar real-world situation – Apply simple calculations (*e.g.,* subtraction, division by two-digit number) to solve problems in a geometric setting
1	*87% of all students across the OECD area can perform tasks at least at Level 1 on the space and shape scale*	
	Solve simple problems in a familiar context using familiar pictures or drawings of geometric objects and applying counting or basic calculation skills	– Use a given two-dimensional representation to count or calculate elements of a simple three-dimensional object

This level has been chosen to align country performance in Figure 2.6a as it represents a baseline level of mathematics proficiency on the PISA scale at which students begin to demonstrate the kind of literacy skills that enable them to actively use mathematics as stipulated by the PISA definition: at Level 2, students demonstrate the use of direct inference to recognise the mathematical elements of a situation, are able to use a single representation to help explore and understand a situation, can use basic algorithms, formulae and procedures, and make literal interpretations and apply direct reasoning. In Finland, more than 90 per cent of students perform at or above this threshold.

The great majority of students, 87 per cent, can at least complete the easiest space and shape tasks required to reach Level 1 (Table 2.1a). However, this also varies greatly across countries.

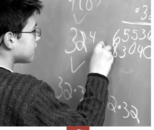
Figure 2.6a ■ **Percentage of students at each level of proficiency on the mathematics/space and shape scale**



Countries are ranked in descending order of percentage of 15-year-olds in Levels 2, 3, 4, 5 and 6.
Source: OECD PISA 2003 database, Table 2.1a.

One way to summarise student performance and to compare the relative standing of countries on the mathematics/space and shape scale is by way of their mean scores. This is shown in Figure 2.6b. As discussed in Box 2.1, when interpreting mean performance, only those differences between countries that are statistically significant should be taken into account. The figure shows those pairs of countries where the difference in their mean scores is sufficient to say with confidence that the higher performance by sampled students in one country holds for the entire population of enrolled 15-year-olds. A country's performance relative to that of the countries listed along the top of the figure can be seen by reading across each row. The colours indicate whether the average performance of the country in the row is either lower than that of the comparison country, not statistically significant different, or higher. When making multiple comparisons, *e.g.,* when comparing the performance of one country with that of all other countries,

An overall mean score of country performance can be compared, but in some cases country differences are not statistically significant…



Box 2.1 ■ Interpreting sample statistics

Standard errors and confidence intervals. The statistics in this report represent *estimates* of national performance based on samples of students rather than the values that could be calculated if every student in every country had answered every question. Consequently, it is important to know the degree of uncertainty inherent in the estimates. In PISA 2003, each estimate has an associated degree of uncertainty, which is expressed through a standard error. The use of confidence intervals provides a means of making inferences about the population means and proportions in a manner that reflects the uncertainty associated with sample estimates. Under the usually reasonable assumption of a normal distribution, and unless otherwise noted in this report, there is a 95 per cent chance that the true value lies within the confidence interval.

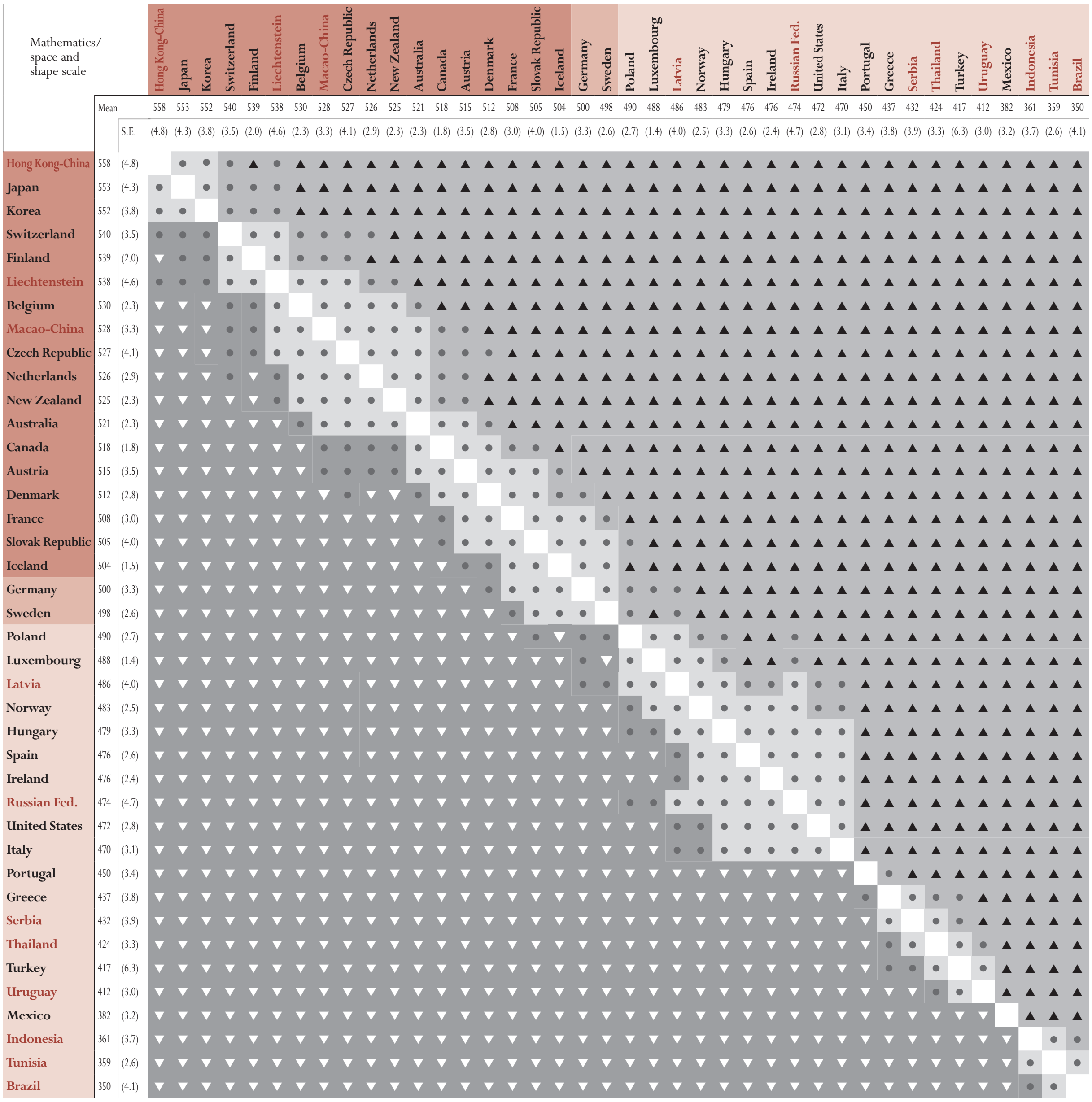
Judging whether populations differ. This report tests the statistical significance of differences between the national samples in percentages and in average performance scores in order to judge whether there are differences between the populations that the samples represent. Each separate test follows the convention that, if in fact there is no real difference between two populations, there is no more than a 5 per cent probability that an observed difference between the two samples will erroneously suggest that the populations are different as the result of sampling and measurement error. In the figures and tables showing multiple comparisons of countries' mean scores, multiple comparison significance tests are also employed that limit to 5 per cent the probability that the mean of a given country will erroneously be declared to be different from that of any other country, in cases where there is in fact no difference (Annex A4).

a more cautious approach is required: only those comparisons indicated by the upward and downward pointing symbols should be considered statistically significant for the purpose of multiple comparisons.[9] Figure 2.6b also shows which countries perform above, at or below the OECD average. Results from the United Kingdom were excluded from this and similar comparisons, because the data for England did not comply with the response rate standards which OECD countries had established to ensure that PISA yields reliable and internationally comparable data (Annex A3).

…so one can only say within a range where each country ranks, with Hong Kong-China, Japan and Korea performing strongest.

For the reasons explained in Box 2.1 it is not possible to determine the exact rank order position of countries in the international comparisons. However, Figure 2.6b shows the range of rank order positions within which the country mean lies with 95 per cent probability. Results are shown both for the OECD countries and all countries that participated in PISA 2003, including both OECD and partner countries. For example, while the mean score for the partner country Hong Kong-China is the highest on the mathematics/space and shape scale followed by the scores from Japan and Korea, it is important to note that they are not statistically different from each other. Because of sampling errors, it is not possible to say which country's rank lies first, but it is possible to say with 95 per cent confidence that Japan, Korea and Hong Kong-China lie between first and third positions of all countries.

Figure 2.6b ■ Multiple comparisons of mean performance on the mathematics/space and shape scale



Range of rank*

		Hong Kong-China	Japan	Korea	Switzerland	Finland	Liechtenstein	Belgium	Macao-China	Czech Republic	Netherlands	New Zealand	Australia	Canada	Austria	Denmark	France	Slovak Republic	Iceland	Germany	Sweden
OECD countries	Upper rank		1	1	3	3		5		5	5	5	7	9	9	10	11	12	13	14	15
	Lower rank		2	2	4	4		7		9	9	9	11	11	13	14	15	17	16	17	17
All countries	Upper rank	1	1	1	4	4	4	6	6	6	7	8	10	12	12	13	14	15	16	17	18
	Lower rank	3	3	3	6	6	8	10	12	12	12	12	14	14	16	17	18	20	19	20	20

		Poland	Luxembourg	Latvia	Norway	Hungary	Spain	Ireland	Russian Fed.	United States	Italy	Portugal	Greece	Serbia	Thailand	Turkey	Uruguay	Mexico	Indonesia	Tunisia	Brazil
OECD countries	Upper rank	17	18		19	20	21	21		22	22	26	27			28		29			
	Lower rank	20	20		21	24	25	24		25	25	26	27			28		29			
All countries	Upper rank	20	21	21	22	23	25	25	24	26	27	31	32	32	33	34	35	37	38	38	39
	Lower rank	23	23	26	26	29	29	29	30	30	30	31	33	34	35	36	36	37	39	40	40

* Because data are based on samples, it is not possible to report exact rank order positions for countries. However, it is possible to report the range of rank order positions within which the country mean lies with 95 per cent likelihood.

Instructions:

Read across the row for a country to compare performance with the countries listed along the top of the chart. The symbols indicate whether the average performance of the country in the row is lower than that of the comparison country, higher than that of the comparison country, or if there is no statistically significant difference between the average achievement of the two countries.

Without the Bonferroni adjustment:
Mean performance statistically significantly higher than in comparison country
No statistically significant difference from comparison country
Mean performance statistically significantly lower than in comparison country

With the Bonferroni adjustment:
▲ Mean performance statistically significantly higher than in comparison country
● No statistically significant difference from comparison country
▽ Mean performance statistically significantly lower than in comparison country

Statistically significantly above the OECD average
Not statistically significantly different from the OECD average
Statistically significantly below the OECD average

Source: OECD PISA 2003 database.

Box 2.2 ■ Interpreting differences in PISA scores: how large a gap?

What is meant by a difference of, say, 50 points between the scores of two different groups of students? The following comparisons can help to judge the magnitude of score differences.

A difference of 62 score points represents one proficiency level on the PISA mathematics scales. This can be considered a comparatively large difference in student performance in substantive terms: for example, with regard to the thinking and reasoning skills that were described above in the section on the process dimension of the PISA 2003 assessment framework, Level 3 requires students to make sequential decisions and to interpret and reason from different information sources, while direct reasoning and literal interpretations are sufficient to succeed at Level 2. Similarly, students at Level 3 need to be able to work with symbolic representations, while for students at Level 2 the handling of basic algorithms, formulae, procedures and conventions is sufficient. With regard to modelling skills, Level 3 requires students to make use of different representational models, while for Level 2 it is sufficient to recognise, apply and interpret basic given models. Students at Level 3 need to use simple problem-solving strategies, while for Level 2 the use of direct inferences is sufficient.

Another benchmark is that the difference in performance on the mathematics scale between the OECD countries with the highest and lowest mean performance is 159 score points, and the performance gap between the countries with the third highest and the third lowest mean performance is 93 score points.

Finally, for the 26 OECD countries in which a sizeable number of 15-year-olds in the PISA samples were enrolled in at least two different grades, the difference between students in the two grades implies that one school year corresponds to an average of 41 score points on the PISA mathematics scale (Table A1.2, Annex A1).[10]

However, since about 90 per cent of performance variation occurs within countries, country averages give only part of the picture.

Finally, it needs to be taken into account that average performance figures mask significant variation in performance within countries, reflecting different levels of performance among many different student groups. As in previous international studies of student performance, such as the IEA Third International Mathematics and Science Study (TIMSS) conducted in 1995 and 1999 and the IEA Trends in Mathematics and Science Study (TIMSS) conducted in 2003, only about one-tenth of the variation in student performance on the overall mathematics scale lies between countries and can, therefore, be captured through a comparison of country averages (Table 5.21a). The remaining variation in student performance occurs within countries, that is, between education systems and programmes, between schools and between students within schools.

In the mathematics/space and shape scale, performance also varies notably between males and females, and more so than in the three other mathematics scales. Gender differences are most clearly visible at the top end of the scale:

on average across countries, 7 per cent of males reach Level 6, while only 4 per cent of females do so and in the Czech Republic, Japan, Korea, the Slovak Republic, Switzerland and the partner country Liechtenstein, the gender gap is around 6 percentage points or larger (Table 2.1b).

Males outperform females in this area of mathematics in most countries, particularly at the top end of the scale.

Nevertheless, in most countries the differences are not large when comparing them over the entire proficiency spectrum.[11] Across the combined OECD area, males perform on average 16 score points higher than females on the mathematics/space and shape scale and they outperform females in all countries except Iceland, where females outperform males. The difference in favour of males reaches more than 35 score points, equivalent to half a proficiency level in mathematics, in the Slovak Republic and in partner country Liechtenstein. However, the overall differences in favour of males are not statistically significant in seven of the participating countries, namely Finland, Japan, the Netherlands and Norway and in the partner countries Hong Kong-China, Serbia and Thailand (Table 2.1c).

Comparison of these results with PISA 2000 must be made with caution…

It is also possible to estimate how much performance on the mathematics/space and shape scale has changed since the last PISA survey in 2000. However, such differences need to be interpreted with caution. Firstly, since data are only available from two points in time, it is not possible to assess to what extent the observed differences are indicative for longer-term trends. Second, while the overall approach to measurement used by PISA is consistent across cycles, small refinements continue to be made, so it would not be prudent to read too much into small changes in results at this stage. Furthermore, sampling and measurement error limit the reliability of comparisons of results over time. Both types of error inevitably arise when assessments are linked through a limited number of common items over time. To account for the effects of such error, the confidence band for comparisons over time has been broadened correspondingly.[12]

…and show little change on average, improvements in four OECD countries and a decline in two.

With these caveats in mind, the following comparisons can be made. On average across OECD countries, performance on the mathematics/space and shape scale has remained broadly similar among the 25 OECD countries for which data can be compared (in 2000, the OECD average was 494 score points whereas in 2003 it was 496 score points). However, when examining performance changes in individual countries, the pattern is uneven (Figures 2.6c and 2.6d, and Table 2.1c and Table 2.1d). In Belgium and Poland, mean performance increases amounted to between 28 and 20 score points, respectively, roughly equivalent to a half grade-year difference in student performance among OECD countries (Box 2.2). The Czech Republic and Italy, as well as the partner countries Brazil, Indonesia, Latvia and Thailand, have also seen significant performance increases in the mathematics/space and shape scale, while performance in Iceland and Mexico declined. In Mexico, this may have been partly attributable to the strong emphasis on increasing participation rates in secondary schools across the country.[13,14] In the remaining countries, there was no statistically significant change in the mean score at the 95 per cent confidence level.

Figure 2.6c ■ **Comparisons between PISA 2003 and PISA 2000 on the mathematics/space and shape scale**

Significance levels	2003 higher than 2000	2003 lower than 2000	No statistically significant difference
90 % confidence level	+	–	O
95 % confidence level	++	– –	
99 % confidence level	+++	– – –	

Differences observed in the mean and percentiles

	5th	10th	25th	Mean	75th	90th	95th
OECD countries							
Australia	O	O	O	O	O	O	O
Austria	O	O	O	O	O	O	O
Belgium	+	O	++	+++	+++	+++	+++
Canada	O	O	O	O	O	O	O
Czech Republic	++	++	++	++	+	O	O
Denmark	---	---	---	-	O	O	O
Finland	++	+	O	O	O	O	O
France	O	O	O	O	O	++	O
Germany	O	O	O	+	+	O	O
Greece	O	O	O	O	--	--	--
Hungary	O	O	O	O	O	+	++
Iceland	---	---	---	--	O	O	O
Ireland	O	O	O	O	O	O	O
Italy	O	O	+	++	++	++	+
Japan	O	O	O	O	O	O	O
Korea	O	O	O	+	O	O	O
Mexico	-	--	--	--	--	-	-
New Zealand	O	O	O	O	O	O	O
Norway	O	O	O	O	O	O	O
Poland	+++	+++	+++	++	O	O	O
Portugal	+++	+++	++	O	O	O	O
Spain	O	O	O	O	O	O	O
Sweden	O	O	O	O	--	--	--
Switzerland	O	O	O	O	O	O	O
United States	O	O	O	O	O	+	+
OECD total	O	O	O	O	O	O	O
OECD average	O	O	O	O	O	O	O
Partner countries							
Brazil	+++	+++	+++	+++	+	O	O
Hong Kong-China	O	O	O	+	+++	+	O
Indonesia	+++	+++	+++	+++	O	O	O
Latvia	+++	+++	+++	+++	++	+	O
Liechtenstein	O	O	O	O	O	O	O
Russian Federation	O	O	O	O	O	O	O
Thailand	+++	+++	++	++	O	O	O

Source: OECD PISA 2003 and PISA 2000 databases, Tables 2.1c and 2.1d.

But it is not just changes in mean scores that are of interest…

Changes in mean performance scores are typically used to assess improvements in the quality of schools and education systems. However, as noted above, mean performance does not provide a full picture of student performance and can mask significant variation within an individual class, school or education system. Moreover, countries aim not only to encourage high performance but also to minimise internal disparities in performance. Both parents and the public at large are aware of the seriousness of low performance and the fact that school-leavers who lack fundamental skills face poor employment prospects. Having a high proportion of students at the lower end of the mathematics scale may give rise to concern that a large proportion of tomorrow's workforce and voters will lack the skills required for the informed judgements that they will need to make.

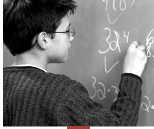
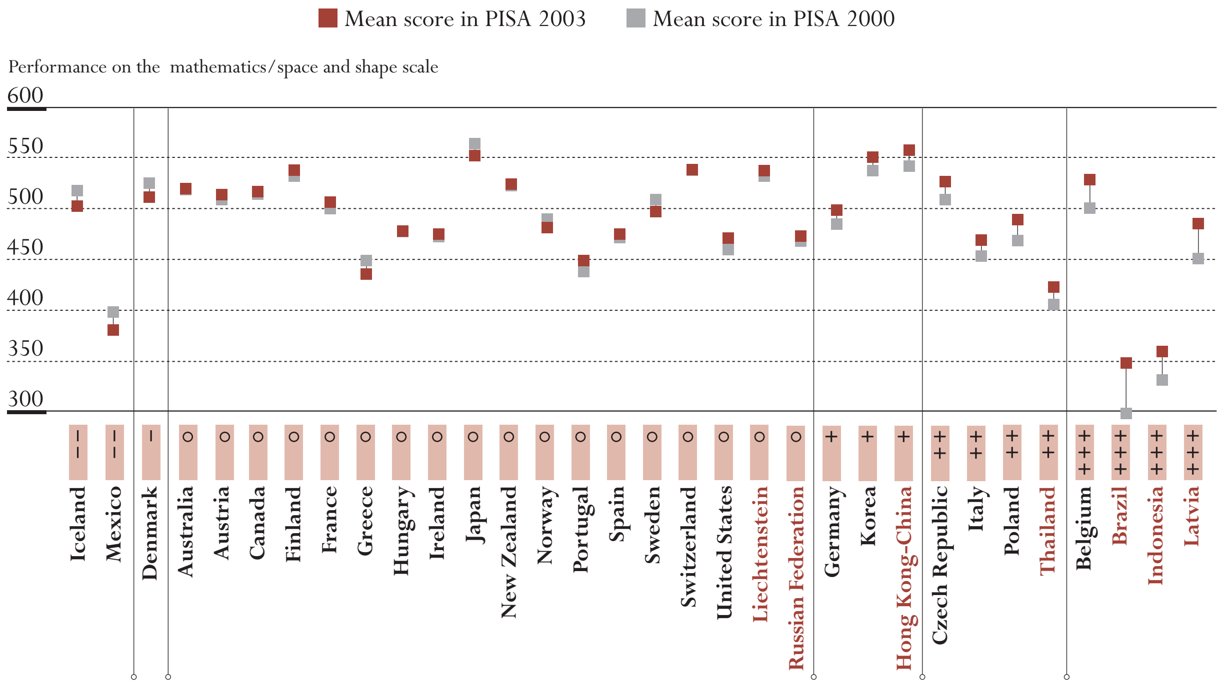


Figure 2.6d ■ **Differences in mean scores between PISA 2003 and PISA 2000 on the mathematics/space and shape scale**

Only countries with valid data for both 2003 and 2000



Significance levels	2003 higher than 2000	2003 lower than 2000	No statistically significant difference
90 % confidence level	+	–	O
95 % confidence level	++	– –	
99 % confidence level	+++	– – –	

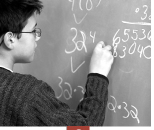
Countries are ranked in ascending order of the difference between PISA 2003 and PISA 2000 performances.
Source: OECD PISA 2003 and PISA 2000 databases, Tables 2.1c and 2.1d.

It is, therefore, important to examine the observed performance changes in more detail. As seen in Figure 2.6c some of the observed changes have not necessarily involved an even rise or fall in performance across the ability range. In some countries, performance across the ability range has widened or narrowed over a three-year period, as changes in one part of the ability range are not matched by changes in others.

…since some change is driven by a particular part of the ability range.

In Belgium, for example, the 28 point rise in average performance on the mathematics/space and shape scale has mainly been driven by improved performance in the top part of the performance distribution – as is visible in the increase in scores at the 75th, 90th and 95th percentiles – while little has changed at the lower end of the distribution (Figures 2.6c and 2.6d, and Tables 2.1c and 2.1d). A similar picture, though less pronounced, emerges for Italy. As a result, in these two countries overall performance increased but the gap between the better and poorer performers has widened.

Improvements in Belgium and Italy have been driven by higher-ability students…



...whereas in Poland and the Czech Republic overall performance increased because lower-performing students tended to catch up.

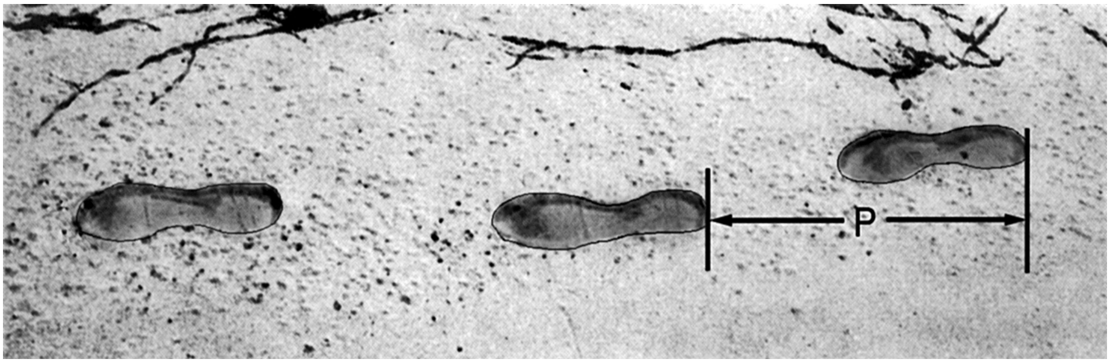
In contrast, for Poland, the rise in average performance on the mathematics/ space and shape scale is attributable mainly to an increase in performance at the lower end of the performance distribution (*i.e.,* 5th, 10th and 25th percentiles). Consequently, in 2003 fewer than 5 per cent of students fell below performance standards that had not been reached by the bottom 10 per cent of Polish students in 2000. As a result, Poland succeeded in raising the average performance of 15-year-olds on the mathematics/space and shape scale while narrowing the overall performance gap between the lower and higher achievers over this period; this change that may well be associated with the massive reform of the schooling systems in 1999, which now provide more integrated educational structures. To a lesser extent, this pattern also holds for the Czech Republic, the remaining country with a substantial increase in average performance (Figures 2.6c-d, Table 2.1c and Table 2.1d).

Student performance on the mathematics/change and relationships scale

A quarter of the mathematical tasks given to students in PISA are related to mathematical manifestations of change, functional relationships and dependency among variables. Figures 2.7a-b show tasks at all six levels in this category:

Figure 2.7a ■ **A sample of mathematics items used in PISA for the change and relationships scale: Unit WALKING**

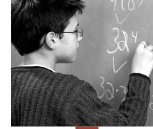
WALKING



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.



WALKING

QUESTION 5

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.

Score 3 (723)

Answers which indicate correctly metres/minute (89.6) and km/hour (5.4). Errors due to rounding are acceptable.

Score 2 (666)

Answers which are incorrect or incomplete because:

- They were not multiplied by 0.80 to convert from steps per minute to metres per minute.
- They correctly showed the speed in metres per minute (89.6 metres per minute) but the conversion to kilometres per hour was incorrect or missing.
- They were based on the correct method (explicitly shown) but with other minor calculation error(s).
- They indicated only 5.4 km/hr, but not 89.6 metres per minute (intermediate calculations not shown).

Score 1 (605)

Answers which give $n = 140 \text{ x } .80 = 112$ but no further working out is shown or incorrect working out from this point.

This open-constructed response item is situated in a personal context. The coding guide for this item provides for full credit, and two levels of partial credit. The item is about the relationship between the number of steps per minute and pacelength. It follows that it fits the change and relationships content area. The mathematical routine needed to solve the problem successfully is substitution in a simple formula (algebra), and carrying out a non-routine calculation. To solve the problem, students first calculate the number of steps per minute when the pace-length is given (0.8 m). This requires substitution into and manipulation of the expression: n/0.8 = 140 leading to: n = 140 x 0.8 which is 112 steps per minute. The next question asks for the speed in m/minute which involves converting the number of steps to a distance in metres: 112 x 0.80 = 89.6 metres; so his speed is 89.6 m/minute. The final step is to transform this speed into km/h - a more commonly used unit of speed. This involves relationships among units for conversions which is part of the measurement domain. Solving the problem also requires decoding and interpreting basic symbolic language, and handling expressions containing symbols and formulae. The problem, therefore, is rather a complex one involving formal algebraic expression and performing a sequence of different but connected calculations that need understanding of transforming formulas and units of measures. The lower level partial credit part of this item belongs to the connections competency cluster and with a difficulty of 605 score points it illustrates the top part of Level 4. The higher level of partial credit illustrates the upper part of Level 5, with a difficulty of 666 score points. Students who score the higher level of partial credit are able to go beyond finding the number of steps per minute, making progress towards converting this into the more standard units of speed asked for. However, their responses are either not entirely complete or not fully correct. Full credit for this item illustrates the upper part of Level 6, as it has a difficulty of 723 score points. Students who score full credit are able to complete the conversions and provide a correct answer in both of the requested units.

QUESTION 4

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

Score 1 (611)

Answers which indicate p = 0.5 m or p = 50 cm or p = ½ (unit not required).

This open-constructed response item is situated in a personal context. It has a difficulty of 611 score points, just 4 points beyond the boundary with Level 4. Everyone has seen his or her own footsteps printed in the sand at some moment in life, most likely without realising what kind of relations exist in the way these patterns are formed, although many students will have an intuitive feeling that if the pace-length increases, the number of steps per minute will decrease, other things equal. To reflect on and realise the embedded mathematics in such daily phenomena is part of acquiring mathematical literacy. The item is about this relationship: number of steps per minute and pacelength. It follows that it fits the change and relationships content area. The mathematical content could be described as belonging clearly to algebra. Students need to solve the problem successfully by substitution in a simple formula and carrying out a routine calculation: if n/p = 140, and n = 70, what is the value of p? The students need to carry out the actual calculation in order to get full credit. The competencies needed involve reproduction of practised knowledge, the performance of routine procedures, application of standard technical skills, manipulation of expressions containing symbols and formulae in standard form, and carrying out computations. Therefore the item belongs to the reproduction competency cluster. The item requires problem solving by making use of a formal algebraic expression. With this combination of competencies, and the real-world setting that students must handle, it illustrates Level 5, at the lower end.

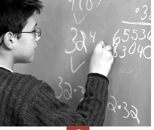
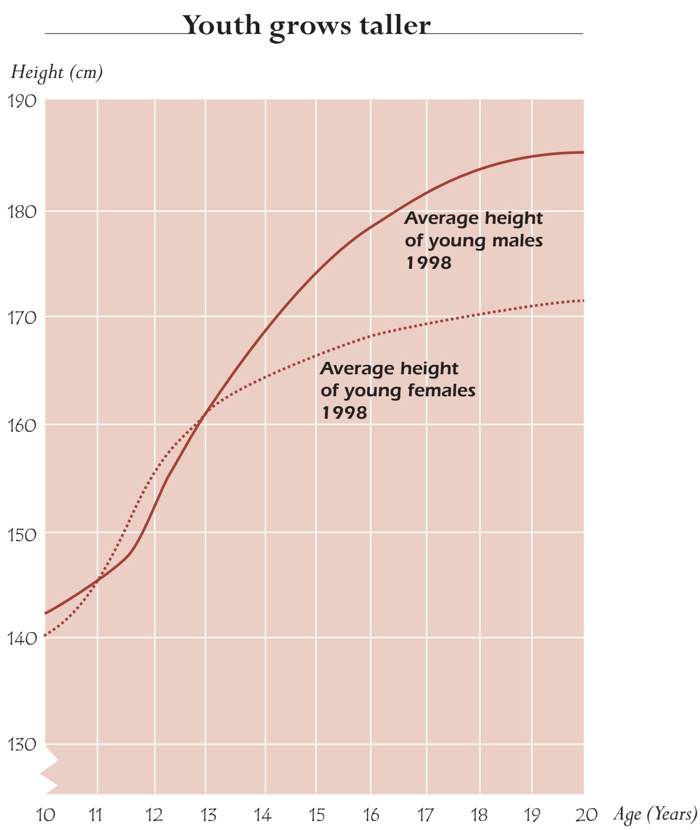
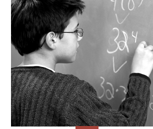


Figure 2.7b ■ **A sample of mathematics items used in PISA for the change and relationships scale: Unit GROWING UP**

GROWING UP

In 1998 the average height of both young males and young females in the Netherlands is represented in this graph.





GROWING UP

QUESTION 8

Explain how the graph shows that on average the growth rate for girls slows down after 12 years of age.

Score 1 (574)

Correct answers which refer to the "change" of the gradient of the graph for females, either by explicitly referring to the reduced steepness of the curve from 12 years onwards, using daily-life or mathematical language, or implicitly by using the actual amount of growth before 12 years and after 12 years of age.

This open-constructed response item has a difficulty of 574 score points (Level 4). The focus of the item is on the relationship between age and height, which means that it belongs to the change and relationships content area. Solving the problem involves the interpretation and decoding of familiar representations of well known mathematical objects. But there is a rather complex concept in this item, the concept of "decreasing growth", which is a combination of "growing" and "slowing down", to use the language of the item. In mathematical terms: the graphs become less steep and the slope (or gradient) decreases. The graphs indicate that this diminished growth rate starts at around age 12. The communication of this observation is central to the question for the students. The expression of their answers ranges from daily life language to more mathematical language about the reduced steepness, or they compare the actual growth in centimetres per year. Thus the mathematical content can be described as evaluating the characteristics of a data set represented in a graph, and noting and interpreting the different slopes at various points of the graphs. In competency terms, the item represents a situation that is not routine but involves familiar settings and demands the linking of different ideas and information – it therefore belongs to the connections competencies cluster. The item requires mathematical insight and some reasoning and communication of the results of this process, within the explicit models of growth.

QUESTION 7

According to this graph, on average, during which period in their life are females taller than males of the same age?

Score 2 (525)

Answers which indicate the correct interval, from 11-13 years or state that girls are taller than boys when they are 11 and 12 years old (this answer is correct in daily-life language, because it means the interval from 11 to 13).

Score 1 (420)

Other subsets of (11, 12, 13), not included in the full credit section.

This item, with its focus on age and height means that it lies in the change and relationships content area - it has a difficulty of 420 (Level 1). The mathematical content can be described as belonging to the data domain because the students are asked to compare characteristics of two data sets, interpret these data sets and draw conclusions. The competencies needed to successfully solve the problem are in the reproduction cluster and involve the interpretation and decoding of reasonably familiar and standard representations of well known mathematical objects. Students need thinking and reasoning competencies to answer the question: "Where do the graphs have common points?" and the argumentation and communication competencies to explain the role these points play in finding the desired answer. Students who score partial credit are able to show that their reasoning and/or insight was well directed, but they fail in coming up with a full, comprehensive answer. They properly identify ages like 11 and/or 12 and/or 13 as being part of an answer but fail to identify the continuum from 11 to 13 years. The item provides a good illustration of the boundary between Level 1 and Level 2. The full credit response to this item illustrates Level 3, as it has a difficulty of 525 score points. Students who score full credit are not only able to show that their reasoning and/or insight is well directed, but they also come up with a full, comprehensive answer. Students who solve the problem successfully are adept at using graphical representations, making conclusions and communicating their findings.

QUESTION 6

Since 1980 the average height of 20-year-old females has increased by 2.3 cm, to 170.6 cm. What was the average height of a 20-year-old female in 1980?

Answer: .... cm

Score 1 (477)

Answers which indicate 168.3 cm (unit already given).

This closed-constructed response item is situated in a scientific context: the growth curves of young males and females over a period of ten years. It has a difficulty of 477 score points. Science uses graphical representation frequently, for example as in this item to represent changes in height in relation to the age. Because of the focus on these aspects this item is classified as belonging to the change and relationships area. The mathematics content is basic. Translating the question into a mathematical context and carrying out a basic arithmetic operation: subtraction (170.6 – 2.3). This places it in the reproduction competency cluster: the thinking and reasoning required involves the most basic form of questions ("How much is the difference?"); the same holds for the argumentation competency: the students just need to follow a standard quantitative process. An added complexity is the fact that the answer can be found by ignoring the graph altogether – an example of redundant information. Summarising, the item requires that students can extract the relevant information from a single source (and ignoring the redundant source) and make use of a single representational mode and can employ a basic subtraction algorithm. Therefore the item illustrates Level 2.

LeveL

6

668.7

5

606.6

4

544.4

3

482.4

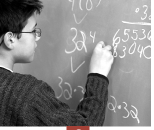
2

420.4

1

358.3

Below 1



A small minority of students can perform the very hardest change and relationships tasks…

The precise competencies required to reach each level are given in Figure 2.8. As with the mathematics/space and shape scale, 5 per cent of students in the combined OECD area can perform Level 6 tasks. Thirty-two per cent of students in the OECD area, but half of the students in Korea, the Netherlands, and the partner country Hong Kong-China, and just under half of the students in Belgium, Finland and the partner country Liechtenstein, and Finland, reach at least Level 4.

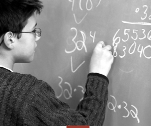
Figure 2.8 ■ **Summary descriptions of six levels of proficiency on the mathematics/change and relationships scale**

Level	General competencies students should have at each level	Specific tasks students should be able to do
6	*5% of all students across the OECD area can perform tasks at Level 6 on the change and relationships scale*	
	Use significant insight, abstract reasoning and argumentation skills and technical knowledge and conventions to solve problems and to generalise mathematical solutions to complex real-world problems	– Interpret complex mathematical information in the context of an unfamiliar real-world situation – Interpret periodic functions in a real-world setting, perform related calculations in the presence of constraints – Interpret complex information hidden in the context of an unfamiliar real-world situation – Interpret complex text and use abstract reasoning (based on insight into relationships) to solve problems – Insightful use of algebra or graphs to solve problems; ability to manipulate algebraic expressions to match a real-world situation – Problem solving based on complex proportional reasoning – Multi-step problem-solving strategies involving the use of formulae and calculations – Devise a strategy and solve the problem by using algebra or trial-and-error – Identify a formula which describes a complex real-world situation, generalise exploratory findings to create a summarising formula – Generalise exploratory findings in order to carry out calculations – Apply deep geometrical insight to work with and generalise complex patterns – Conceptualise complex percentage calculations – Coherently communicate logical reasoning and arguments
5	*15% of all students across the OECD area can perform tasks at least at Level 5 on the change and relationships scale*	
	Solve problems by making advanced use of algebraic and other formal mathematical expressions and models; link formal mathematical representations to complex real-world situations; use complex and multi-step problem-solving skills, reflect on and communicate reasoning and arguments	– Interpret complex formulae in a scientific context – Interpret periodic functions in a real-world setting, perform related calculations – Use advanced problem-solving strategies – Interpret and link complex information – Interpret and apply constraints – Identify and carry out a suitable strategy – Reflect on the relationship between an algebraic formula and its underlying data – Use complex proportional reasoning, *e.g.,* related to rates – Analyse and apply a given formula in a real-life situation – Communicate reasoning and argument
4	*32% of all students across the OECD area can perform tasks at least at Level 4 on the change and relationships scale*	
	Understand and work with multiple representations, including explicit mathematical models of real-world situations to solve practical problems; employ considerable flexibility in interpretation and reasoning, including in unfamiliar contexts, and communicate the resulting explanations and arguments	– Interpret complex graphs, and read one or multiple values from graphs – Interpret complex and unfamiliar graphical representations of real-world situations – Use multiple representations to solve a practical problem – Relate text-based information to a graphic representation and communicate explanations – Analyse a formula describing a real-world situation – Analyse three-dimensional geometric situations involving volume and related functions

Level	General competencies students should have at each level	Specific tasks students should be able to do
		– Analyse a given mathematical model involving a complex formula – Interpret and apply word formulae, and manipulate and use linear formulae that represent real-world relationships – Carry out a sequence of calculations involving percentages, proportions, addition or division
3	*54% of all students across the OECD area can perform tasks at least at Level 3 on the change and relationships scale*	
	Solve problems that involve working with multiple related representations (a text, a graph, a table, a formula), including some interpretation, reasoning in familiar contexts, and communication of argument	– Interpret unfamiliar graphical representations of real-world situations – Identify relevant criteria in a text – Interpret text in which a simple algorithm is hidden and apply that algorithm – Interpret a text and devise a simple strategy – Link and connect multiple related representations (*e.g.,* two related graphs, text and a table, a formula and a graph) – Use reasoning involving proportions in various familiar contexts and communicate reasons and argument – Apply a text-given criterion or situation to a graph – Use a range of simple calculation procedures to solve problems, including ordering data, time difference calculations and linear interpolation
2	*73% of all students across the OECD area can perform tasks at least at Level 2 on the change and relationships scale*	
	Work with simple algorithms, formulae and procedures to solve problems; link text with a single representation (a graph, a table, a simple formula); use interpretation and reasoning skills at an elementary level	– Interpret a simple text and link it correctly to graphical elements – Interpret a simple text that describes a simple algorithm and apply that algorithm – Interpret a simple text and use proportional reasoning or a calculation – Interpret a simple pattern – Interpret and use reasoning in a practical context involving a simple and familiar application of motion, speed and time relationships – Locate relevant information in graph, and read values directly from a graph – Correctly substitute numbers to apply a simple numeric algorithm or simple algebraic formula
1	*87% of all students across the OECD area can perform tasks at least at Level 1 on the change and relationships scale*	
	Locate relevant information in a simple table or graph; follow direct and simple instructions to read information directly from a simple table or graph in a standard or familiar form; perform simple calculations involving relationships between two familiar variables	– Make a simple connection of text to a specific feature of a simple graph and read off a value from the graph – Locate and read a specified value in a simple table – Perform simple calculations involving relationships between two familiar variables

Seventy-three per cent of students in the combined OECD area perform at least at Level 2, the level that was chosen to align the results in Figure 2.9a. It represents, as explained above, a baseline level of mathematics proficiency on the PISA scale at which students begin to demonstrate the kind of literacy skills that enable them to actively use mathematics as stipulated by the PISA definition (Table 2.2a). However in Greece, Italy, Luxembourg, Mexico, Norway, Poland, Portugal, Spain, Turkey and the United States as well as in the partner countries Brazil, Indonesia, Latvia, the Russian Federation, Serbia, Thailand, Tunisia and Uruguay a quarter or more of students fail to reach this threshold.

…and about one in four cannot perform more than the very simplest tasks.



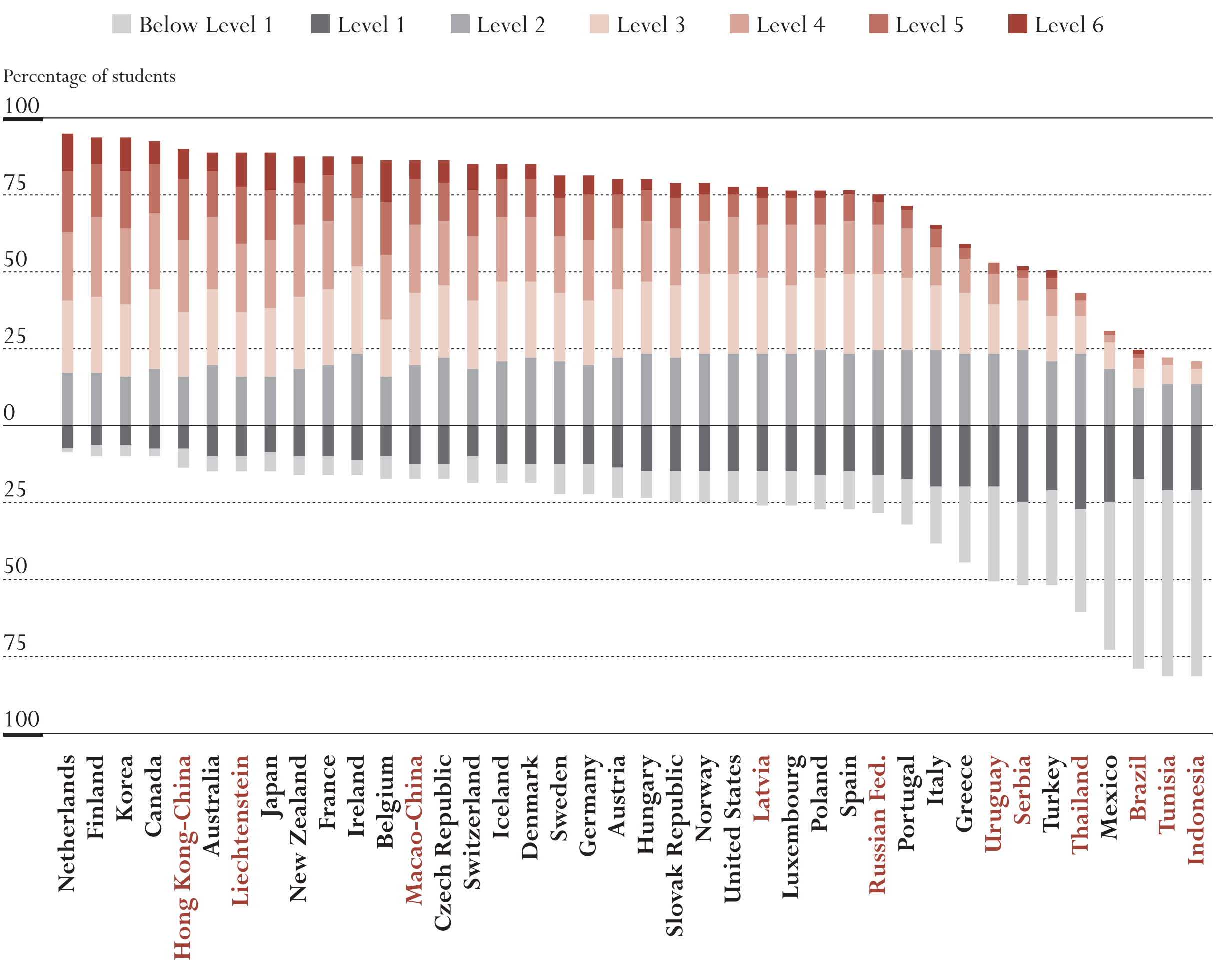
There is a larger country gap on this mathematics scale than in any other…

Among the various mathematics scales, the change and relationships scale shows the largest gap in mean performance between high and low performing countries – 214 score points or more separate the Netherlands at half a standard deviation above the OECD average from Brazil, Indonesia and Tunisia at more than one and a half standard deviations below the OECD average (Figure 2.9b).

…and again, the overall performance can be compared across countries, with the Netherlands, Finland, Korea and Hong Kong-China the strongest.

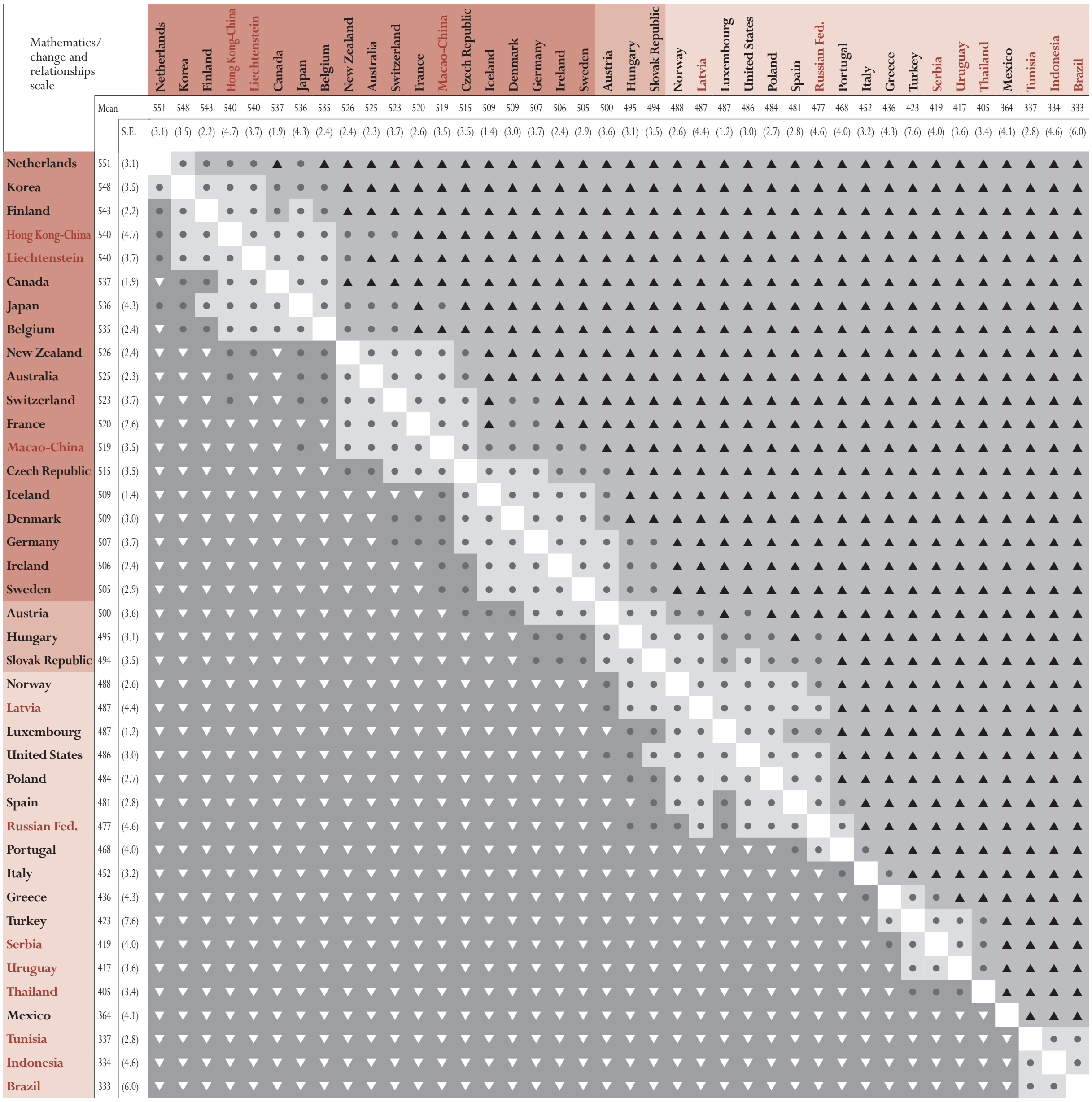
Figure 2.9b gives a summary of overall student performance in different countries on the change and relationships scale, in terms of the mean student score, and shows, with 95 per cent probability, the range of rank order positions within which the country mean lies. As explained before, it is not possible to determine the exact rank order position of countries in the international comparisons. However, it can be concluded that the Netherlands' position is between first and third among all countries that participated in PISA 2003, indistinguishable from Korea which can be found between the first and fourth ranks.

Figure 2.9a ■ **Percentage of students at each level of proficiency on the mathematics/change and relationships scale**



Countries are ranked in descending order of percentage of 15-year-olds in Levels 2, 3, 4, 5 and 6.
Source: OECD PISA 2003 database, Table 2.2a.

Figure 2.9b ■ Multiple comparisons of mean performance on the mathematics/change and relationships scale



Range of rank*

		Netherlands	Korea	Finland	Hong Kong-China	Liechtenstein	Canada	Japan	Belgium	New Zealand	Australia	Switzerland	France	Macao-China	Czech Republic	Iceland	Denmark	Germany	Ireland	Sweden	Austria
OECD countries	Upper rank	1	1	2			4	3	4	7	7	7	8		10	11	11	11	12	12	14
	Lower rank	2	3	4			6	7	6	9	10	11	11		14	15	16	17	17	17	19
All countries	Upper rank	1	1	2	2	2	4	3	5	8	9	9	10	10	12	14	14	14	15	15	17
	Lower rank	3	4	6	8	8	8	9	8	12	12	13	14	14	17	18	19	20	20	20	22

		Hungary	Slovak Republic	Norway	Latvia	Luxembourg	United States	Poland	Spain	Russian Fed.	Portugal	Italy	Greece	Turkey	Serbia	Uruguay	Thailand	Mexico	Tunisia	Indonesia	Brazil
OECD countries	Upper rank	17	17	19		20	20	20	21		25	26	27	27				29			
	Lower rank	20	21	23		23	24	24	24		25	26	28	28				29			
All countries	Upper rank	20	20	22	21	23	23	23	25	26	29	31	32	32	33	33	35	37	38	38	38
	Lower rank	23	24	27	28	27	28	29	29	30	30	31	33	35	35	35	36	37	40	40	40

* Because data are based on samples, it is not possible to report exact rank order positions for countries. However, it is possible to report the range of rank order positions within which the country mean lies with 95 per cent likelihood.

Instructions:

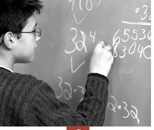
Read across the row for a country to compare performance with the countries listed along the top of the chart. The symbols indicate whether the average performance of the country in the row is lower than that of the comparison country, higher than that of the comparison country, or if there is no statistically significant difference between the average achievement of the two countries.

Without the Bonferroni adjustment:
- Mean performance statistically significantly higher than in comparison country
- No statistically significant difference from comparison country
- Mean performance statistically significantly lower than in comparison country

With the Bonferroni adjustment:
- ▲ Mean performance statistically significantly higher than in comparison country
- ● No statistically significant difference from comparison country
- ▽ Mean performance statistically significantly lower than in comparison country

- Statistically significantly above the OECD average
- Not statistically significantly different from the OECD average
- Statistically significantly below the OECD average

Source: OECD PISA 2003 database.



Males outperform females in just over half of the countries.

Males outperform females in 17 OECD countries and four partner countries, but generally only by small amounts (Table 2.2c).[15] The average performance difference between males and females is only 10 score points, that is, a somewhat smaller gap than the difference found for the mathematics/space and shape scale. Only in Iceland do females perform higher than males. Nevertheless, as in the case of the mathematics/space and shape scale, gender differences tend to be larger at the top end of the scale (Table 2.2b).

Results on this scale can also be compared, with caution, to PISA 2000…

As for the mathematics/space and shape scale, it is also possible to estimate how much performance has changed since PISA 2000 (Table 2.2c and Table 2.2d). However, as explained in the preceding section, these differences need to be interpreted with caution since data are only available from two points in time, while the observed differences are not only influenced by sampling error but are also subject to the uncertainty associated with the linking of the two assessments.

…showing that performance in change and relationships tasks rose overall, but unevenly…

On average across OECD countries, performance among the 25 countries for which data can be compared has increased from 488 score points in 2000 to 499 score points in 2003, the biggest overall change observed in any area of the PISA assessment. But again, changes have been very uneven across OECD countries. The Czech Republic and Poland and the partner countries Brazil, Latvia and Liechtenstein have seen increases of 31 to 70 score points in mean performance – equivalent to between half and one PISA proficiency level – and in Belgium, Canada, Finland, Germany, Hungary, Korea, Portugal and Spain increases were still between 13 and 22 score points. For the remaining countries, the differences cannot be considered statistically significant when both measurement and assessment linkage errors are taken into account.[16]

..again driven in some countries by improvements among lower ability students…

As with the mathematics/space and shape scale, some of the observed changes have not necessarily involved an even rise or fall of performance across the ability range (Figures 2.9c and 2.9d). The large improvements in Poland have been driven by improved performance at the lower end of the performance distribution (*i.e.,* 5th, 10th and 25th percentiles). As a result, Poland succeeded in significantly raising the average performance of 15-year-olds in the mathematics/change and relationships scale and narrowing the overall performance gap between the lower and higher achievers over this period. A similar picture, though less pronounced, is also evident in the Czech Republic and Hungary as well as in the partner countries Latvia and Liechtenstein. Also Greece and Switzerland as well as in the partner country the Russian Federation have seen notable improvements at the lower end of the distribution, but these were not sufficient to lead to a statistically significant improvement in mean performance.

…but for others by higher ability students.

In contrast, in Canada, Finland, Germany, Italy, Korea, Portugal and Sweden, improvements in performance have mainly been driven by improved performance in the top part of the performance distribution, as shown in the increase in scores at the 75th, 90th and 95th percentiles, while less has changed at

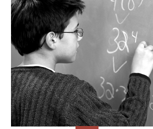


Figure 2.9c ■ **Comparisons between PISA 2003 and PISA 2000 on the mathematics/change and relationships scale**

Significance levels	2003 higher than 2000	2003 lower than 2000	No statistically significant difference
90 % confidence level	+	–	O
95 % confidence level	++	– –	
99 % confidence level	+++	– – –	

Differences observed in the mean and percentiles

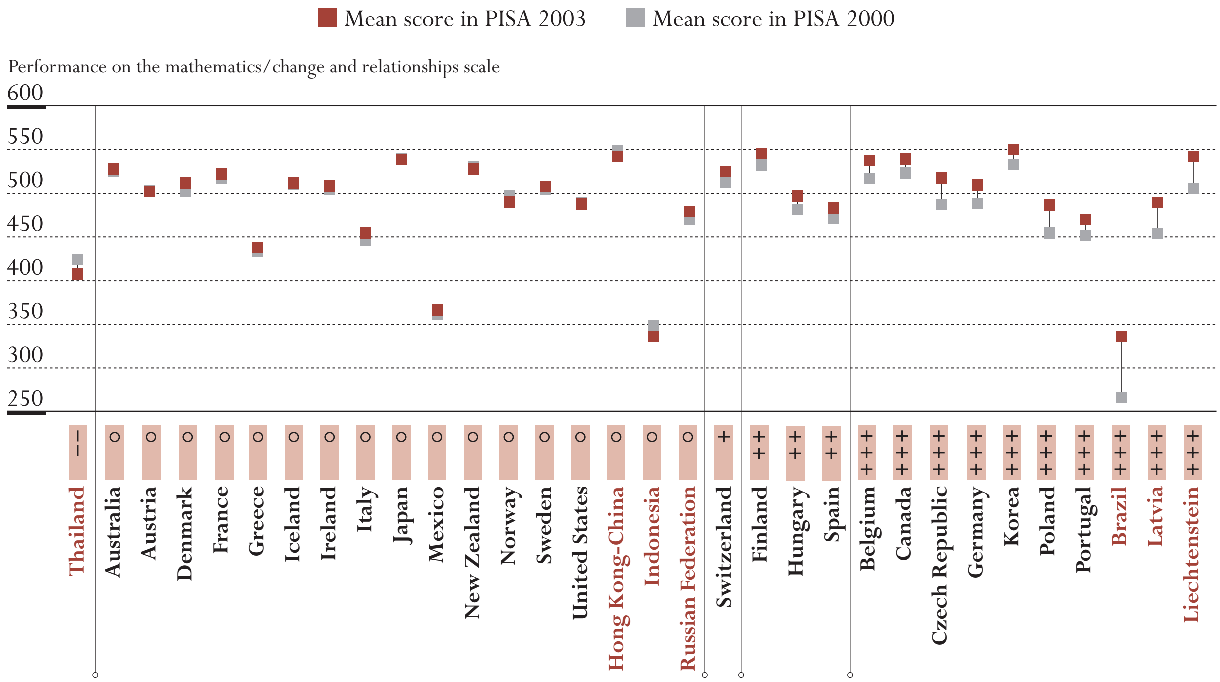
	5th	10th	25th	Mean	75th	90th	95th
OECD countries							
Australia	O	O	O	O	O	O	O
Austria	O	O	O	O	O	O	O
Belgium	+++	+	+	+++	+++	+++	+
Canada	++	++	++	+++	+++	+++	+++
Czech Republic	+++	+++	+++	+++	+++	++	+
Denmark	++	+	O	O	O	O	O
Finland	O	+	O	++	+++	+++	+++
France	O	O	O	O	O	O	O
Germany	++	+	++	+++	+++	+++	+++
Greece	+++	++	O	O	O	-	---
Hungary	+++	+++	+++	++	O	O	O
Iceland	O	O	O	O	O	O	O
Ireland	O	O	O	O	O	+	O
Italy	O	O	O	O	O	++	+++
Japan	O	O	O	O	O	O	O
Korea	O	O	O	+++	+++	+++	+++
Mexico	O	O	O	O	O	O	O
New Zealand	O	O	O	O	O	O	O
Norway	O	O	O	O	O	O	O
Poland	+++	+++	+++	+++	O	O	O
Portugal	+	+	+	+++	+++	++	+++
Spain	+	+	++	++	+	O	O
Sweden	O	O	O	O	O	++	++
Switzerland	+++	+++	++	+	O	O	O
United States	O	O	O	O	O	O	O
OECD total	O	O	O	O	O	O	O
OECD average	+++	+++	+	++	++	++	++
Partner countries							
Brazil	+++	+++	+++	+++	+++	+++	+++
Hong Kong-China	O	-	O	O	O	O	O
Indonesia	---	---	---	O	O	+++	+++
Latvia	+++	+++	+++	+++	+	O	O
Liechtenstein	++	++	+++	+++	O	O	O
Russian Federation	+++	+++	++	O	O	-	-
Thailand	---	---	---	--	O	+++	+++

Source: OECD PISA 2003 and PISA 2000 databases, Tables 2.2c and 2.2d.

the lower end of the distribution. In some of these countries, disparities among students have grown. In the 2000 assessment, for example, Korea showed the smallest variation in student performance in mathematics. By contrast, in the 2003 assessment variation is now at the OECD average level (Figure 2.9c, Figure 2.9d, Table 2.2c and Table 2.2d).

Figure 2.9d ■ Differences in mean scores between PISA 2003 and PISA 2000 on the mathematics/change and relationships scale

Only countries with valid data for both PISA 2003 and PISA 2000



Significance levels	2003 higher than 2000	2003 lower than 2000	No statistically significant difference
90 % confidence level	+	–	O
95 % confidence level	++	– –	
99 % confidence level	+++	– – –	

Countries are ranked in ascending order of the difference between PISA 2003 and PISA 2000 performances.
Source: OECD PISA 2003 and PISA 2000 databases, Tables 2.2c and 2.2d.

Student performance on the mathematics/quantity scale

A quarter of the mathematical tasks given to students in PISA related to numeric phenomena and quantitative relationships and patterns. Figures 2.10a-b show tasks at Levels 1-4 in this category:

Four per cent of students in the OECD area can perform the hardest quantity tasks…

The precise competencies required to reach each level are explained in Figure 2.11. Slightly fewer students than for the previous two scales, at 4 per cent in the combined OECD area, can perform at Level 6 tasks. Slightly more, at 74 per cent, can perform at Level 2 (Table 2.3a). However, in Greece, Italy, Mexico, Portugal, Turkey and the United States, as well as in the partner countries Brazil, Indonesia, the Russian Federation, Serbia, Thailand, Tunisia and Uruguay, a quarter or more of students fail to reach this Level 2 threshold (Figure 2.12a).

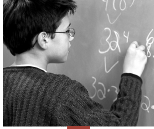


Figure 2.10a ■ **A sample of mathematics items used in PISA for the quantity scale: Unit EXCHANGE RATE**

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 11

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.

Score 1 (586)

Answers which indicate 'Yes', with adequate explanation.

This open-constructed response item is situated in a public context and has a difficulty of 586 score points. As far as the mathematics content is concerned students need to apply procedural knowledge involving number operations: multiplication and division, which along with the quantitative context, places the item in the quantity area. The competencies needed to solve the problem are not trivial: students need to reflect on the concept of exchange rate and its consequences in this particular situation. The mathematisation required is of a rather high level although all the required information is explicitly presented: not only is the identification of the relevant mathematics somewhat complex, but also the reduction to a problem within the mathematical world places significant demands on the student. The competency needed to solve this problem can be described as using flexible reasoning and reflection. The thinking and reasoning competency, the argumentation competency in combination with the problem-solving competency all include an element of reflectiveness on the part of the student about the process needed to solve the problem. Explaining the results requires some communication skills as well. Therefore the item is classified as belonging to the reflection cluster. The combination of familiar context, complex situation, non-routine problem, the need for reasoning and insight and a communication demand places the item in Level 4.

QUESTION 10

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?

Score 1 (439)

Answers which indicate 975 SGD (unit not required).

This short-constructed response item is situated in a public context. It has a difficulty of 439 score points. The mathematics content is restricted to a basic operation: division. This places the item in the quantity area, and more specifically: operations with numbers. Regarding the competencies required, a limited form of mathematisation is needed: understanding a simple text, in which all the required information is explicitly presented. But students also need to recognise that division is the right procedure to go with, which makes it less trivial than Exchange Rate Question 1, and shows the most basic form of the thinking and reasoning competency. Thus the competency needed to solve this problem can be described as performance of a routine procedure and/or application of a standard algorithm. Therefore the item is classified as belonging the reproduction competency cluster. The combination of familiar context, clearly defined question, and rather routine procedure that includes some decision-making places the item in Level 2.

QUESTION 9

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?

Score 1 (406)

Answers which indicate 12 600 ZAR (unit not required).

This short constructed response item is situated in a public context. It has a difficulty of 406 score points. Experience in using exchange rates may not be common to all students, but the concept can be seen as belonging to skills and knowledge for intelligent citizenship. The mathematics content is restricted to one of the four basic operations: multiplication. This places the item in the quantity area, and more specifically: operations with numbers. As far as the competencies are concerned, a very limited form of mathematisation is needed: understanding a simple text, and linking the given information to the required calculation. All the required information is explicitly presented. Thus the competency needed to solve this problem can be described as performance of a routine procedure and/or application of a standard algorithm. Therefore the item is classified as belonging the reproduction competency cluster. The combination of familiar context, clearly defined question, and routine procedure places the item in Level 1.

Level
6
668.7
5
606.6
4
544.4
3
482.4
2
420.4
1
358.3
Below 1

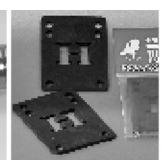
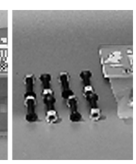
Figure 2.10b ■ **A sample of mathematics items used in PISA for the quantity scale: Unit SKATEBOARD**

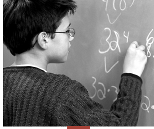
SKATEBOARD

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	  



SKATEBOARD

QUESTION 13

The shop offers three different decks, two different sets of wheels and two different sets of hardware. There is only one choice for a set of trucks.

How many different skateboards can Eric construct?

A. 6
B. 8
C. 10
D. 12

Score 1 (570)

The correct answer is option D.

This multiple-choice item is situated in the personal context and has a difficulty of 570 score points (Level 4). All the required information in this item is explicitly presented and the mathematics involves the basic routine computation: 3 x 2 x 2 x 1. However, if students do not have experience with such combinatorial calculations, their strategy might involve a systematic listing of the possible combinations. There are well-known algorithms for this (such as a tree diagram). The strategy to find the number of combinations can be considered as common, and routine. It involves following and justifying standard quantitative processes, including computational processes, statements and results. Therefore, the item can be classified as belonging to the reproduction competency cluster. The computation involved fits in the quantity content area. In order to be successful the students have to accurately apply an algorithm, after correctly interpreting text in combination with a table. This adds to the complexity of the situation.

QUESTION 14

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	

Score 1 (554)

Answers which indicate 65 zeds on a deck, 14 on wheels, 16 on trucks and 20 on hardware.

This short constructed response item is also in the personal context. and illustrates the lower part of Level 4, (554 score points) The item fits in the quantity content area as the students are asked to compute what is the most expensive skateboard you can buy for 120 zeds. The task, however, is not straightforward as there is no standard procedure or routine algorithm available. As far as the competencies needed, the problem solving skill here involves a more independent approach and students may use different strategies in order to find the solution, including trial and error. The setting of this problem can be regarded as familiar. Students have to look at the table with prices, make combinations and do some computation. This places the item within the connections competency cluster. A strategy that will work with this problem is to first use all the higher values, and then adjust the answer, working the way down until the desired maximum of 120 zeds is reached. Thus, students need some reasoning skills in a familiar context, they have to connect the question with the data given in the table, apply a non-standard strategy and carry out routine calculations.

QUESTION 12

Eric wants to assemble his own skateboard. What is the minimum price and the maximum price in this shop for self-assembled skateboards?

(a) Minimum price: ....zeds.
(b) Maximum price: ....zeds.

Score 2 (496)

Answers which indicate both the minimum (80) and the maximum (137) prices.

Score 1 (464)

Answers which indicate only the minimum (80) or the maximum (137) prices.

This short constructed response item is in a personal context because skateboards tend to be part of the youth culture. The students are asked to find a minimum and maximum price for the construction of a skateboard. The partial credit response has a difficulty of 464 score points (Level 2) - this is when the students answer the question by giving either the minimum or the maximum, but not both. To solve the problem the students have to find a strategy, which is fairly simple because the strategy that seems trivial actually works: the minimum cost is based on the lower numbers and the maximum, on the larger numbers. The remaining mathematics content is execution of a basic operation. The addition: 40 + 14 + 16 + 10 = 80, gives the minimum, while the maximum is found by adding the larger numbers: 65 + 36 + 16 + 20 = 137. The strategy, therefore, is the reproduction of practised knowledge in combination with the performance of the routine addition procedure - this item belongs to the reproduction competency cluster and the quantity content area. The full credit response, when students give both the minimum and the maximum, has a difficulty of 496 score points and illustrates Level 3.

Level
6
668.7
5
606.6
4
544.4
3
482.4
2
420.4
1
358.3
Below 1

Figure 2.11 ■ **Summary descriptions of six levels of proficiency on the mathematics/quantity scale**

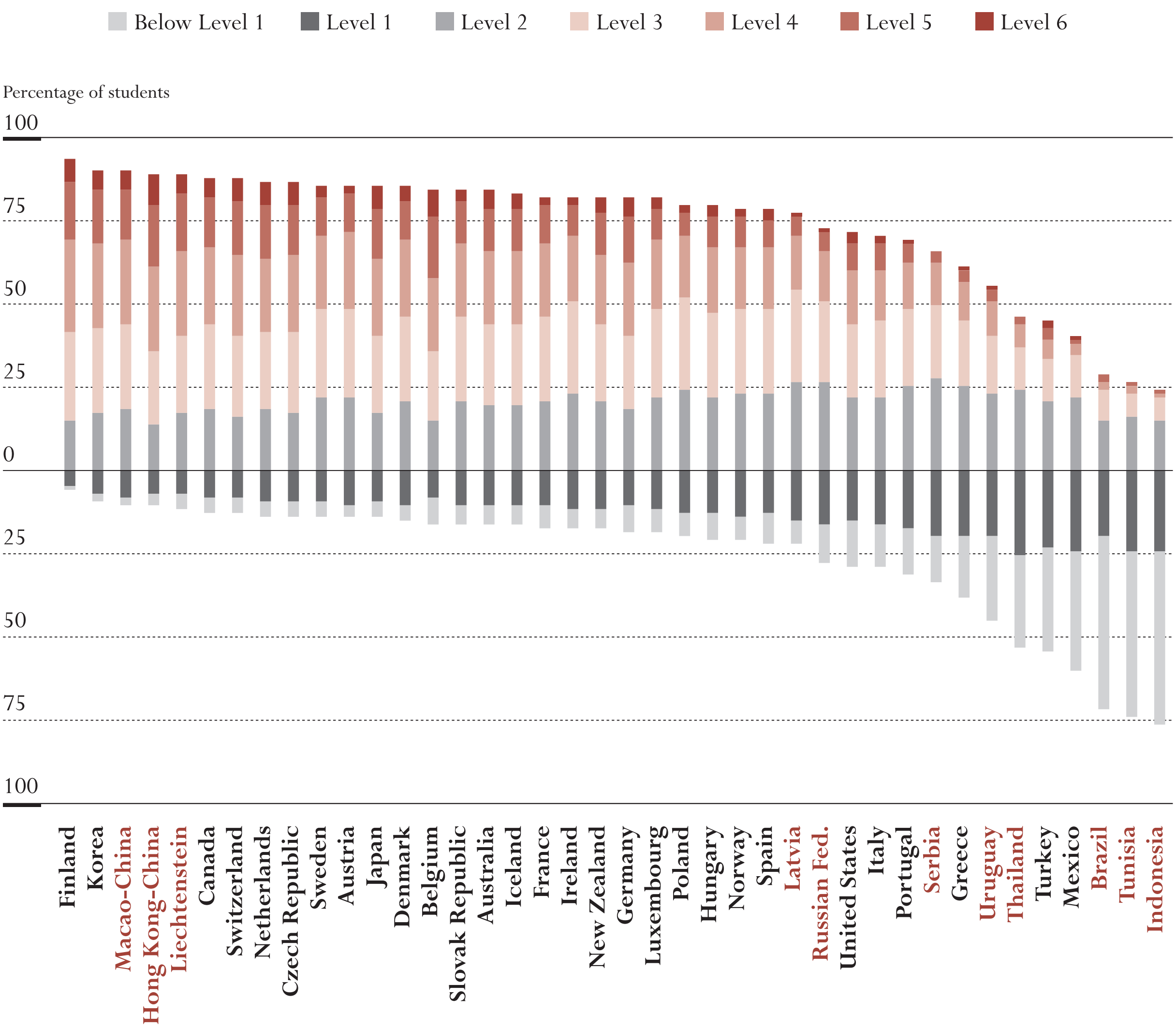
Level	General competencies students should have at each level	Specific tasks students should be able to do
6	*4% of all students across the OECD area can perform tasks at Level 6 on the quantity scale*	
	Conceptualise and work with models of complex mathematical processes and relationships; work with formal and symbolic expressions; use advanced reasoning skills to devise strategies for solving problems and to link multiple contexts; use sequential calculation processes; formulate conclusions, arguments and precise explanations	– Conceptualise complex mathematical processes such as exponential growth, weighted average, as well as number properties and numeric relationships – Interpret and understand complex information, and link multiple complex information sources – Use advanced reasoning concerning proportions, geometric representations of quantities, combinatorics and integer number relationships – Interpret and understand formal mathematical expressions of relationships among numbers, including in a scientific context – Perform sequential calculations in a complex and unfamiliar context, including working with large numbers – Formulate conclusions, arguments and precise explanations – Devise a strategy (develop heuristics) for working with complex mathematical processes
5	*13% of all students across the OECD area can perform tasks at least at Level 5 on the quantity scale*	
	Work effectively with models of more complex situations to solve problems; use well-developed reasoning skills, insight and interpretation with different representations; carry out sequential processes; communicate reasoning and argument	– Interpret complex information about real-world situations (including graphs, drawings and complex tables) – Link different information sources (such as graphs, tabular data and related text) – Extract relevant data from a description of a complex situation and perform calculations – Use problem-solving skills (*e.g.*, interpretation, devising a strategy, reasoning, systematic counting) in real-world contexts that involve substantial mathematisation – Communicate reasoning and argument – Make an estimation using daily life knowledge – Calculate relative and/or absolute change
4	*31% of all students across the OECD area can perform tasks at least at Level 4 on the quantity scale*	
	Work effectively with simple models of complex situations; use reasoning skills in a variety of contexts, interpret different representations of the same situation; analyse and apply quantitative relationships; use a variety of calculation skills to solve problems	– Accurately apply a given numeric algorithm involving a number of steps – Interpret complex text descriptions of a sequential process – Relate text-based information to a graphic representation – Perform calculations involving proportional reasoning, divisibility or percentages in simple models of complex situations – Perform systematic listing and counting of combinatorial outcomes – Identify and use information from multiple sources – Analyse and apply a simple system – Interpret complex text to produce a simple mathematical model

Level	General competencies students should have at each level	Specific tasks students should be able to do
3	*53% of all students across the OECD area can perform tasks at least at Level 3 on the quantity scale*	
	Use simple problem-solving strategies including reasoning in familiar contexts; interpret tables to locate information; carry out explicitly described calculations including sequential processes	– Interpret a text description of a sequential calculation process, and correctly implement the process – Use basic problem-solving processes (devise a simple strategy, look for relationships, understand and work with given constraints, use trial and error, simple reasoning) – Perform calculations including working with large numbers, calculations with speed and time, conversion of units (*e.g.,* from annual rate to daily rate) – Interpret tabular information, locate relevant data from a table – Conceptualise relationships involving circular motion and time – Interpret text and diagrams describing a simple pattern
2	*74% of all students across the OECD area can perform tasks at least at Level 2 on the quantity scale*	
	Interpret simple tables to identify and extract relevant information; carry out basic arithmetic calculations; interpret and work with simple quantitative relationships	– Interpret a simple quantitative model (*e.g.,* a proportional relationship) and apply it using basic arithmetic calculations – Interpret simple tabular data, link textual information to related tabular data – Identify the simple calculation required to solve a straight-forward problem – Perform simple calculations involving the basic arithmetic operations, as well as ordering numbers
1	*88% of all students across the OECD area can perform tasks at least at Level 1 on the quantity scale*	
	Solve problems of the most basic type in which all relevant information is explicitly presented, the situation is straight forward and very limited in scope, the required computational activity is obvious and the mathematical task is basic, such as a simple arithmetic operation	– Interpret a simple, explicit mathematical relationship, and apply it directly using a calculation – Read and interpret a simple table of numbers, total the columns and compare the results

Figure 2.12b gives a summary of overall student performance in different countries on the quantity scale, in terms of mean student scores as well as the range of rank order positions within which the country mean lies with 95 per cent probability. Finland shows the highest mean score among OECD countries on the mathematics/quantity scale but the partner country Hong Kong-China performs at a similarly high level, within the range of the first and third position.

...in which Finland and Hong Kong-China show the highest performance.

Figure 2.12a ■ **Percentage of students at each level of proficiency on the mathematics/quantity scale**



Countries are ranked in descending order of percentage of 15-year-olds in Levels 2, 3, 4, 5 and 6.
Source: OECD PISA 2003 database, Table 2.3a.

In these tasks males' advantage is particularly small.

Consistent with what was found in the other scales, males show an advantage also in the quantity scale, but gender differences here tend to be even smaller than for the mathematics/space and shape and change and relationships scales discussed above. The distributions of males and females by level are relatively similar, with a few more males than females at the top end of the scale (Table 2.3b). Sixteen countries show differences in favour of males.[17] Again, Iceland is the only country where females perform statistically above males (Table 2.3c).

It is not possible to compare student performance in 2000 and 2003 on this scale, since the PISA 2000 assessment did not include this content in its assessment.

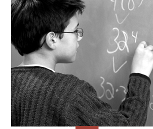
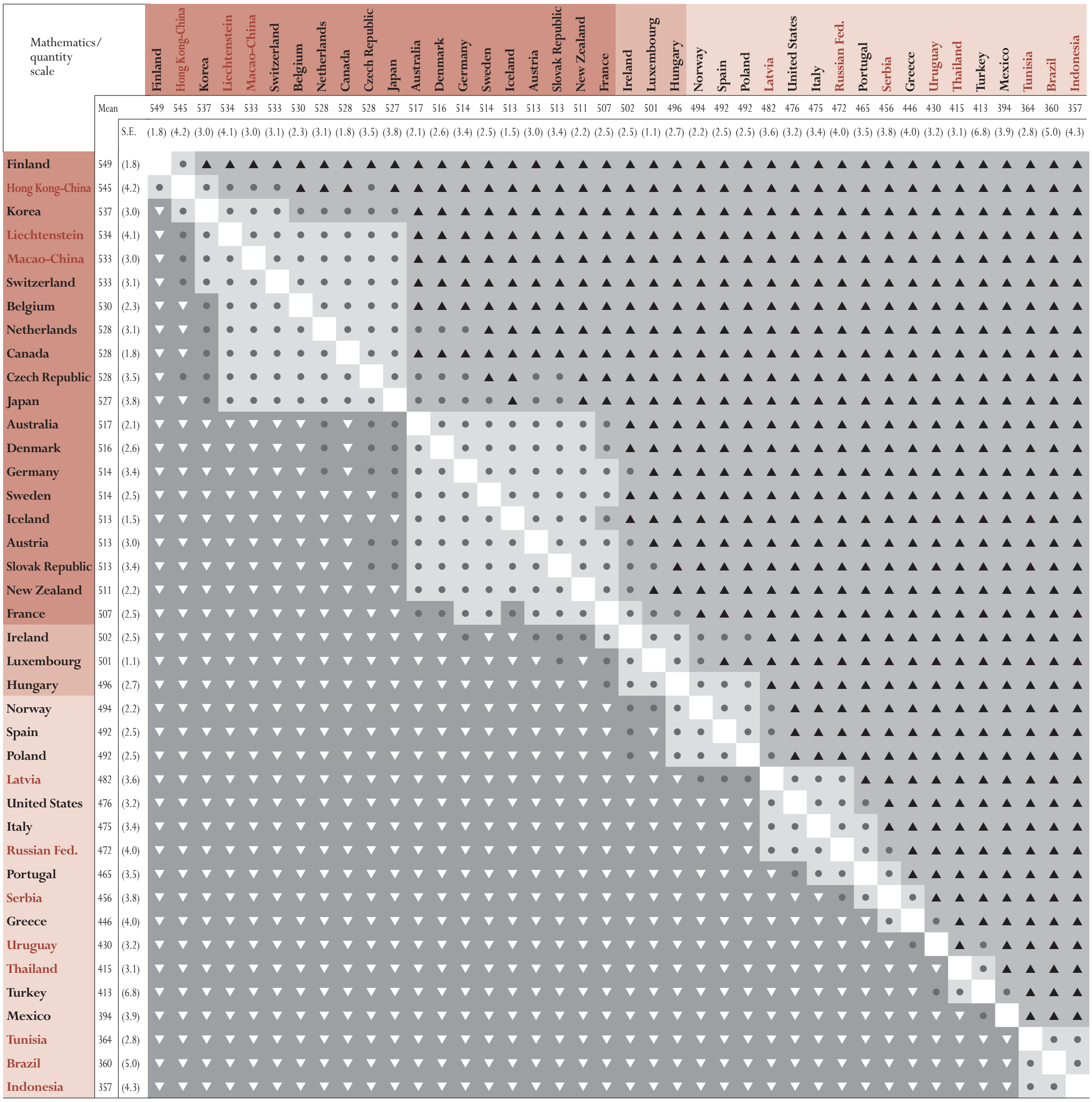


Figure 2.12b ■ Multiple comparisons of mean performance on the mathematics/quantity scale



Range of rank*

		Finland	Hong Kong-China	Korea	Liechtenstein	Macao-China	Switzerland	Belgium	Netherlands	Canada	Czech Republic	Japan	Australia	Denmark	Germany	Sweden	Iceland	Austria	Slovak Republic	New Zealand	France
OECD countries	Upper rank	1		2			2	3	3	4	3	3	9	9	9	9	10	9	9	11	14
	Lower rank	1		4			7	8	8	8	8	8	13	15	16	17	16	17	17	17	18
All countries	Upper rank	1	1	2	3	3	3	4	5	6	4	5	12	12	12	12	13	12	12	14	17
	Lower rank	2	3	7	10	10	10	11	11	11	11	11	16	18	19	20	19	20	20	20	21

		Ireland	Luxembourg	Hungary	Norway	Spain	Poland	Latvia	United States	Italy	Russian Fed.	Portugal	Serbia	Greece	Uruguay	Thailand	Turkey	Mexico	Tunisia	Brazil	Indonesia
OECD countries	Upper rank	17	18	19	20	20	20		24	24		25		27			28	29			
	Lower rank	20	20	23	23	23	23		25	25		26		27			28	29			
All countries	Upper rank	20	21	22	23	23	23	27	27	27	28	30	31	32	34	35	35	37	38	38	38
	Lower rank	23	23	26	26	26	26	29	30	30	31	32	33	33	34	36	36	37	40	40	40

* Because data are based on samples, it is not possible to report exact rank order positions for countries. However, it is possible to report the range of rank order positions within which the country mean lies with 95 per cent likelihood.

Instructions:

Read across the row for a country to compare performance with the countries listed along the top of the chart. The symbols indicate whether the average performance of the country in the row is lower than that of the comparison country, higher than that of the comparison country, or if there is no statistically significant difference between the average achievement of the two countries.

Without the Bonferroni adjustment:
Mean performance statistically significantly higher than in comparison country
No statistically significant difference from comparison country
Mean performance statistically significantly lower than in comparison country

With the Bonferroni adjustment:
▲ Mean performance statistically significantly higher than in comparison country
● No statistically significant difference from comparison country
▽ Mean performance statistically significantly lower than in comparison country

Statistically significantly above the OECD average
Not statistically significantly different from the OECD average
Statistically significantly below the OECD average

Source: OECD PISA 2003 database.

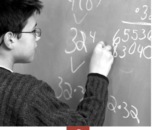
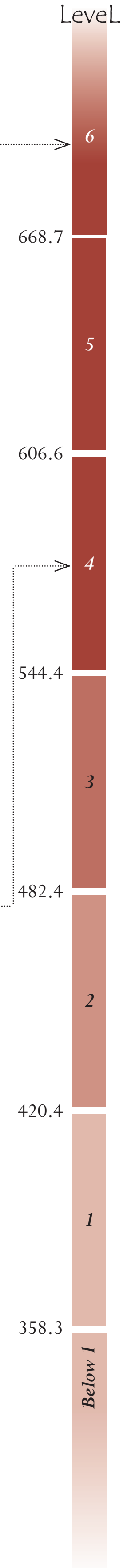


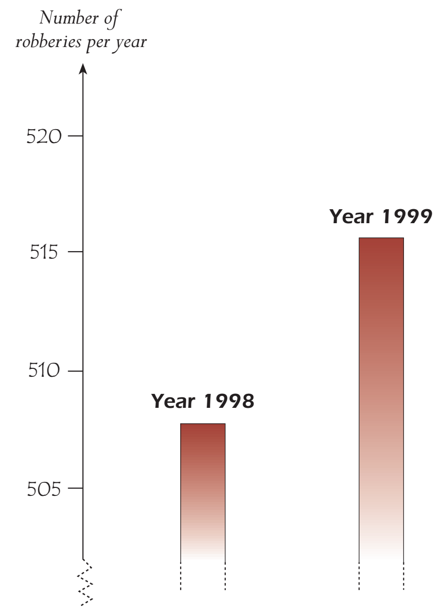
Figure 2.13a ■ **A sample of mathematics items used in PISA for the uncertainty scale: Unit ROBBERIES**



ROBBERIES

A TV reporter showed this graph and said:

"The graph shows that there is a huge increase in the number of robberies from 1998 to 1999."



QUESTION 15

Do you consider the reporter's statement to be a reasonable interpretation of the graph? Give an explanation to support your answer.

Score 2 (694)

Answers which indicate "No, not reasonable" and focus on the fact that only a small part of the graph is shown, or contain correct arguments in terms of ratio or percentage increase, or refer to requirement of trend data before a judgement can be made.

Score 1 (577)

Answers which indicate "No, not reasonable" but explanation lacks detail (focuses ONLY on an increase given by the exact number of robberies, but does not compare with the total) or with correct method but with minor computational errors.

This open-constructed response item is situated in a public context. The graph as presented in the stimulus of this item actually was derived from a real graph with a similarly misleading message as the one here. The graph seems to indicate, as the TV reporter said: "a huge increase in the number of robberies". The students are asked if the statement fits the data. It is very important to look through data and graphs as they are frequently presented in the media in order to participate effectively in society. This constitutes an essential skill in mathematical literacy. Quite often designers of graphics use their skills (or lack thereof) to let the data support a pre-determined message, often with a political context. This is an example. The item involves the analysis of a graph and data interpretation, placing it in the uncertainty area. The reasoning and interpretation competencies required, together with the communication skills needed, are clearly belonging to the connections competency cluster. The competencies that are essential for solving this problem are understanding and decoding of a graphical representation in a critical way, making judgments and finding appropriate argumentation based on mathematical thinking and reasoning (although the graph seems to indicate quite a big jump in the number of robberies, the absolute number of increase in robberies is far from dramatic; the reason for this paradox lies is the inappropriate cut in the y-axis) and proper communication of this reasoning process.

A partial credit response illustrates Level 4 with a difficulty of 577 points. In this case students typically indicate that the statement is not reasonable, but fail to explain their judgment in appropriate detail. This means here that the reasoning only focuses on an increase given by an exact number of robberies in absolute terms, but not in relative terms. Communication is critical here, since one will always have answers that are difficult to interpret in detail. An example: "an increase from 508 to 515 is not large" might have a different meaning from "an increase of around 10 is not large". The first statement shows the actual numbers, and thus the intended meaning of the answer might be that the increase is small because of the large numbers involved, while this line of reasoning does not apply to the second answer. In this kind of response, students use and communicate argumentation based on interpretation of data; therefore it illustrates Level 4.

A full credit response illustrates Level 6 with a difficulty score of 694 score points. In the case of full credit the students indicate that the statement is not reasonable, and explain their judgment in appropriate detail. This means here that the reasoning not only focuses on an increase given by an exact number of robberies in absolute terms, but also in relative terms. The question requires students to use and communicate argumentation based on interpretation of data, using some proportional reasoning in a statistical context, and in a not-too-familiar situation. Therefore it illustrates Level 6.

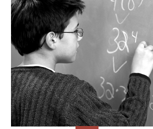
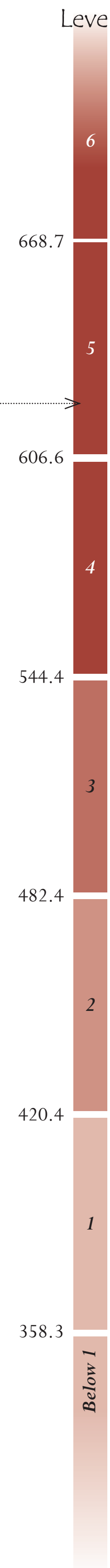


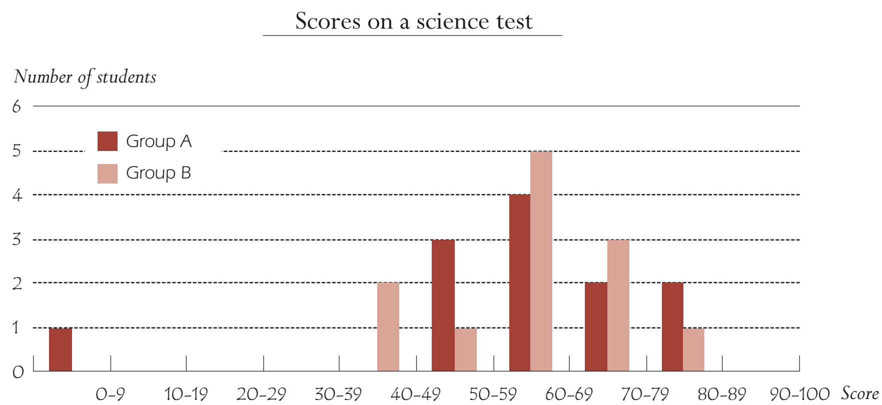
Figure 2.13b ■ **A sample of mathematics items used in PISA for the uncertainty scale: Unit TEST SCORES**



TEST SCORES

The diagram shows the results on a science test for two groups, labelled as Group A and Group B.

The mean score for Group A is 62.0 and the mean for Group B is 64.5. Students pass this test when their score is 50 or above.



QUESTION 16

Looking at the diagram, the teacher claims that Group B did better than Group A in this test.

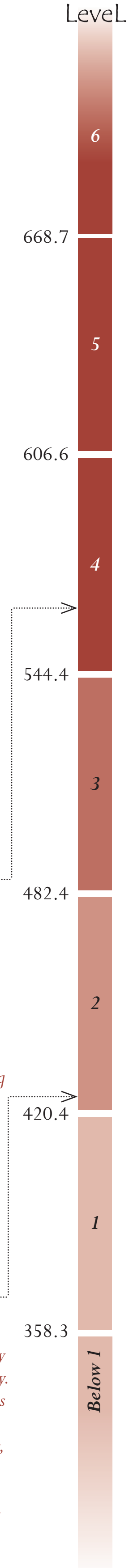
The students in Group A don't agree with their teacher. They try to convince the teacher that Group B may not necessarily have done better.

Give one mathematical argument, using the graph that the students in Group A could use.

Score 1 (620)
Answers which present a valid argument. Valid arguments could relate to the number of students passing, the disproportionate influence of the outlier, or the number of students with scores in the highest level.

This open-constructed response item is situated in an educational context. It has a difficulty of 620 score points. The educational context of this item is one that all students are familiar with: comparing test scores. In this case a science test has been administered to two groups of students: A and B. The results are given to the students in two different ways: in words with some data embedded and by means of two graphs in one grid. The problem is to find arguments that support the statement that Group A actually did better than Group B, given the counter-argument of one teacher that Group B did better – on the grounds of the higher mean for Group B. It will be clear that the item falls into the content area of uncertainty. Knowledge of this area of mathematics is essential in the information society, as data and graphical representations play a major role in the media and in other aspects of our daily experience. The connections cluster, in which this item is classified, includes competencies that not only build on those required for the reproduction cluster (like encoding and interpretation of simple graphical representations) but also require reasoning and insight in a particular mathematical argument. Actually the students have a choice of at least three arguments here. The first one is that more students in Group A pass the test; a second one is the distorting effect of the outlier in the results of Group A; and finally Group A has more students that scored 80 or over. Another important competency needed is explaining matters that include relationships. From this it follows that the item belongs to the connections cluster. Students who are successful have applied statistical knowledge in a problem situation that is somewhat structured and where the mathematical representation is partially apparent. They also need reasoning and insight to interpret and analyse the given information, and they must communicate their reasons and arguments. Therefore the item clearly illustrates Level 5.

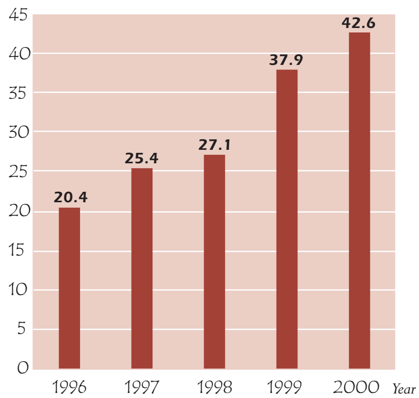
Figure 2.13c ■ **A sample of mathematics items used in PISA for the uncertainty scale: Unit EXPORTS**



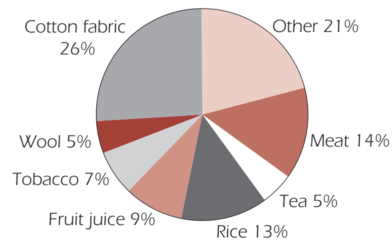
EXPORTS

The graphics 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 18

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.

Score 1 (565)

The correct answer is E. 3.8 million zeds.

This multiple-choice item is situated in a public context. It has a difficulty of 565 score points. The data-handling processes involved with this item place it in the uncertainty area. The mathematical content consists of reading data from two graphs: a bar chart and a pie chart, comparing the characteristics of the two graphics, and combining data from the two graphs in order to be able to carry out a basic number operation resulting in a numerical answer. Students need to combine the information of the two graphics in a relevant way. The mathematisation process needed here has distinct phases: decoding the different standard representations by looking at the total of annual exports of 2000 (42.6 million zeds) and at the percentage of this total coming from fruit juice exports (9%). It is this activity and the process of connecting these numbers by an appropriate numerical operation (9% of 42.6) that places this item in the connections competency cluster. It is the more complex concrete situation, containing two related graphical representations, the insight needed to connect and combine them and the application of the appropriate basic mathematical routine in the relevant way that makes this item fit into Level 4.

QUESTION 17

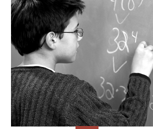
What was the total value (in millions of zeds) of exports from Zedland in 1998?

Answer:

Score 1 (427)

Answers which indicate 27.1 million zeds or 27 100 000 zeds or 27.1 (unit not required). Rounding to 27 also accepted.

This closed-constructed response item is situated in a public context. It has a difficulty of 427 score points. The knowledge society relies heavily on data, and data are often represented in graphics. The media use graphics often to illustrate news articles and make points more convincingly. Reading and understanding this kind of information therefore is an essential component of mathematical literacy. The mathematical content is restricted to reading data from a bar graph or pie chart. Exploratory data analysis is the area of mathematics to which this item belongs, and therefore fits the content area uncertainty. The representation competency is needed to solve this problem: decoding and interpreting a familiar, practised standard representation of a well known mathematical object – following the written instructions, deciding which of the two graphs is relevant and locating the correct information in that graph. This is a routine procedure and therefore the item belongs to the reproduction competency cluster. This item illustrates interpreting and recognising situations in contexts that require no more than direct inference, which is a key feature of Level 2.



Student performance on the mathematics/uncertainty scale

A quarter of the mathematical tasks given to students in PISA related to probabilistic and statistical phenomena and relationships. Figures 2.13a-c shows examples of tasks in Levels 2, 4, 5 and 6 in this category.

Four per cent of students in the OECD area can perform the hardest uncertainty tasks…

The particular competencies required to reach each level are given in Figure 2.14. Only 4 per cent of students in the combined OECD area – but 13 per cent in the partner country Hong Kong-China – can perform Level 6 tasks. Thirty-one per cent of the combined student population in the OECD perform at least at Level 4, but this figure is more than 50 per cent in Finland, the Netherlands and the partner country Hong Kong-China (Table 2.4a).

Figure 2.14 ■ **Summary descriptions of six levels of proficiency on the mathematics/uncertainty scale**

Level	General competencies students should have at each level	Specific tasks students should be able to do
6	*4% of all students across the OECD area can perform tasks at Level 6 on the uncertainty scale*	
	Use high-level thinking and reasoning skills in statistical or probabilistic contexts to create mathematical representations of real-world situations; use insight and reflection to solve problems, and to formulate and communicate arguments and explanations	– Interpret and reflect on real-world situations using probability knowledge and carry out resulting calculations using proportional reasoning, large numbers and rounding – Show insight into probability in a practical context – Use interpretation, logical reasoning and insight at a high level in an unfamiliar probabilistic situation – Use rigorous argumentation based on insightful interpretation of data – Employ complex reasoning using statistical concepts – Show understanding of basic ideas of sampling and carry out calculations with weighted averages, or using insightful systematic counting strategies – Communicate complex arguments and explanations
5	*13% of all students across the OECD area can perform tasks at least at Level 5 on the uncertainty scale*	
	Apply probabilistic and statistical knowledge in problem situations that are somewhat structured and where the mathematical representation is partially apparent. Use reasoning and insight to interpret and analyse given information, to develop appropriate models and to perform sequential calculation processes; communicate reasons and arguments	– Interpret and reflect on the outcomes of an unfamiliar probabilistic experiment – Interpret text using technical language and translate to an appropriate probability calculation – Identify and extract relevant information, and interpret and link information from multiple sources (*e.g.,* from text, multiple tables, graphs) – Use reflection and insight into standard probabilistic situations – Apply probability concepts to analyse a non-familiar phenomenon or situation – Use proportional reasoning and reasoning with statistical concepts – Use multistep reasoning based on data – Carry out complex modelling involving the application of probability knowledge and statistical concepts (*e.g.,* randomness, sample, independence) – Use calculations including addition, proportions, multiplication of large numbers, rounding, to solve problems in non-trivial statistical contexts – Carry out a sequence of related calculations – Carry out and communicate probabilistic reasoning and argument

Level	General competencies students should have at each level	Specific tasks students should be able to do
4	*31% of all students across the OECD area can perform tasks at least at Level 4 on the uncertainty scale*	
	Use basic statistical and probabilistic concepts combined with numerical reasoning in less familiar contexts to solve simple problems; carry out multi-step or sequential calculation processes; use and communicate argumentation based on interpretation of data	– Interpret text, including in an unfamiliar (scientific) but straight-forward context – Show insight into aspects of data from tables and graphs – Translate text description into appropriate probability calculation – Identify and select data from various statistical graphs and carry out basic calculation – Show understanding of basic statistical concepts and definitions (probability, expected value, randomness, average) – Use knowledge of basic probability to solve problems – Construct a basic mathematical explanation of a verbal real-world quantitative concept ("huge increase") – Use mathematical argumentation based on data – Use numerical reasoning – Carry out multi-step calculations involving the basic arithmetic operations, and working with percentage – Draw information from a table and communicate a simple argument based on that information
3	*54% of all students across the OECD area can perform tasks at least at Level 3 on the uncertainty scale*	
	Interpret statistical information and data, and link different information sources; basic reasoning with simple probability concepts, symbols and conventions and communication of reasoning	– Interpret tabular information – Interpret and read from non-standard graphs – Use reasoning to identify probability outcomes in the context of a complex but well-defined and familiar probability experiment – Insight into aspects of data presentation, *e.g.,* number sense; link related information from two different tables; link data to suitable chart type – Communicate common-sense reasoning
2	*75% of all students across the OECD area can perform tasks at least at Level 2 on the uncertainty scale*	
	Locate statistical information presented in familiar graphical form; understand basic statistical concepts and conventions	– Identify relevant information in a simple and familiar graph – Link text to a related graph, in a common and familiar form – Understand and explain simple statistical calculations (*e.g.,* the average) – Read values directly from a familiar data display, such as a bar graph
1	*90% of all students across the OECD area can perform tasks at least at Level 1 on the uncertainty scale*	
	Understand and use basic probabilistic ideas in familiar experimental contexts	– Understand basic probability concepts in the context of a simple and familiar experiment (*e.g.,* involving dice or coins) – Systematic listing and counting of combinatorial outcomes in a limited and well-defined game situation

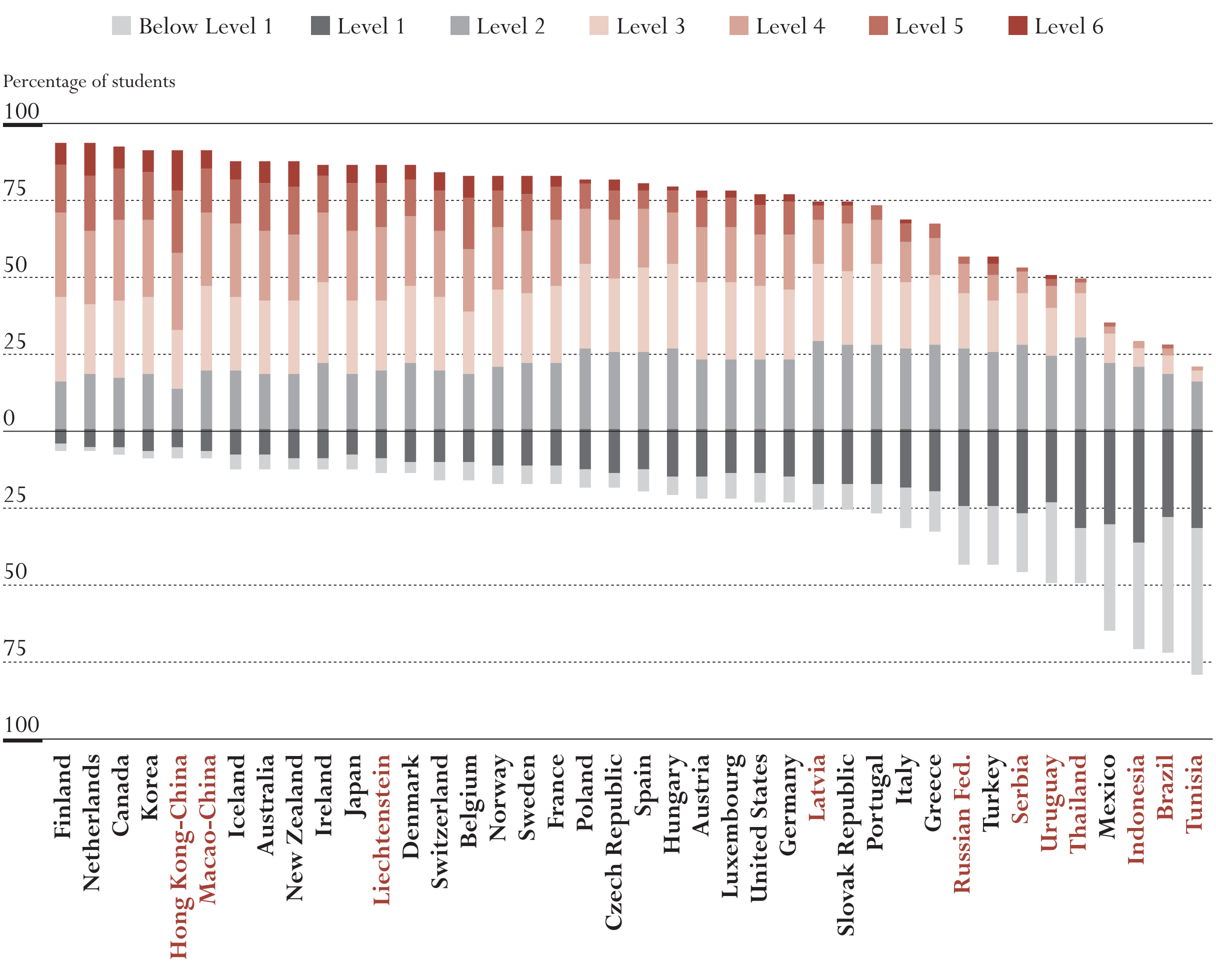
Seventy-five per cent of OECD students can at least function at the baseline Level 2. However, a quarter or more of students fail to reach this threshold in Greece, Italy, Mexico, Portugal, the Slovak Republic and Turkey as well as in the partner countries Brazil, Indonesia, Latvia, the Russian Federation, Serbia, Thailand, Tunisia and Uruguay (Figure 2.15a and Table 2.4a).

…and again a quarter are capable only of the simplest tasks.

Figure 2.15b gives a summary of overall student performance in different countries on the uncertainty scale. Performance is presented in terms of mean student scores as well as, with 95 per cent probability, the range of rank order positions within which the country mean lies. Hong Kong-China and the Netherlands show the strongest performance on the mathematics/uncertainty scale, and can be found between the first and second, and first and third rank order positions, respectively, among all participating countries.

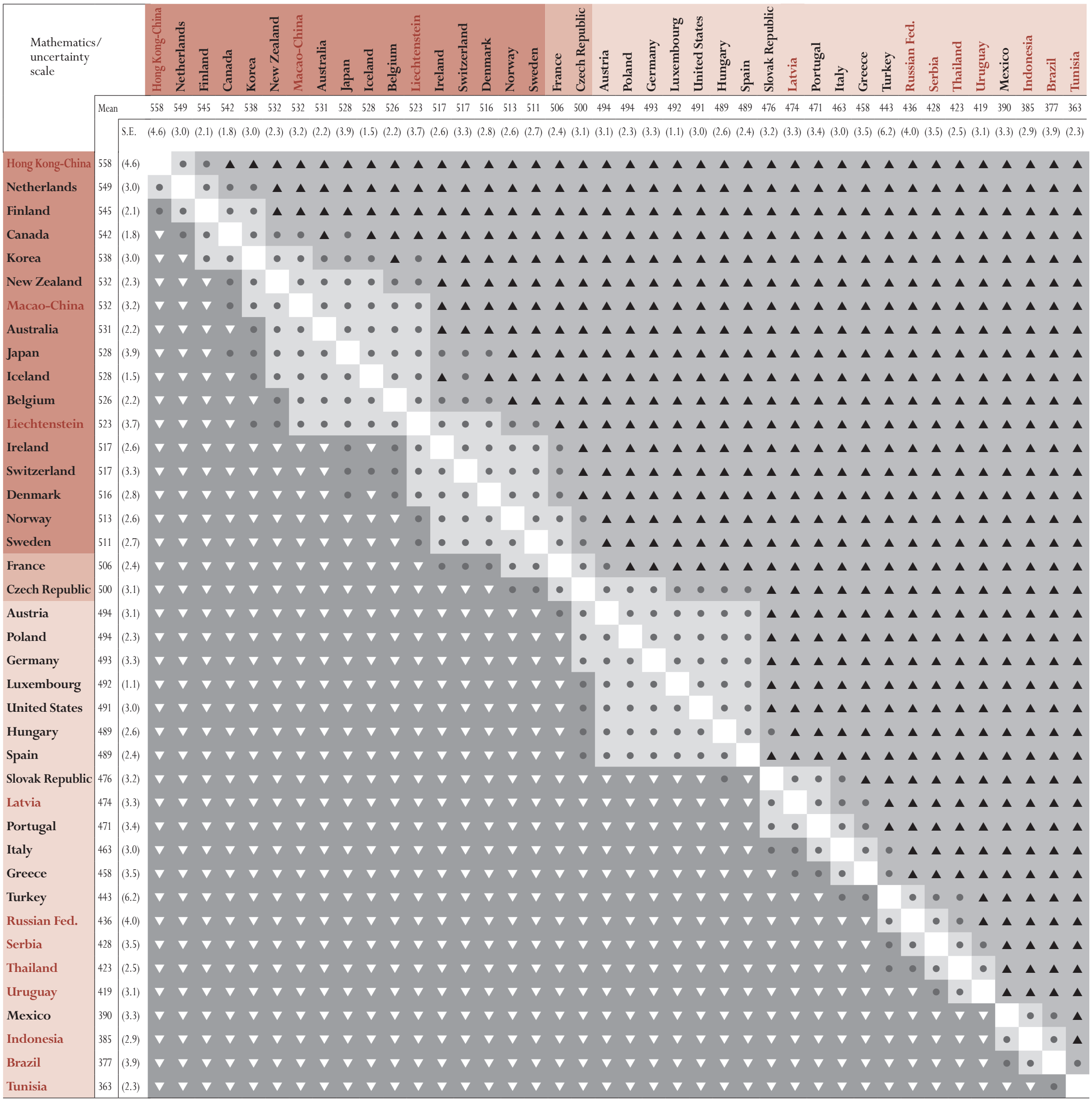
In uncertainty tasks, Hong Kong-China and the Netherlands are strongest overall.

Figure 2.15a ■ **Percentage of students at each level of proficiency on the mathematics/uncertainty scale**



Countries are ranked in descending order of percentage of 15-year-olds in Levels 2, 3, 4, 5 and 6.
Source: OECD PISA 2003 database, Table 2.4a.

Figure 2.15b ■ Multiple comparisons of mean performance on the mathematics/uncertainty scale



Range of rank*

		Hong Kong-China	Netherlands	Finland	Canada	Korea	New Zealand	Macao-China	Australia	Japan	Iceland	Belgium	Liechtenstein	Ireland	Switzerland	Denmark	Norway	Sweden	France	Czech Republic	Austria
OECD countries	Upper rank		1	1	2	2	4		4	5	6	6		10	10	10	10	11	14	15	16
	Lower rank		2	3	4	5	8		8	9	9	9		13	14	14	14	15	16	18	23
All countries	Upper rank	1	1	2	3	4	5	5	6	6	7	8	8	12	12	12	13	14	17	18	19
	Lower rank	2	3	4	5	7	10	11	10	12	11	12	14	16	17	17	17	18	19	21	26

		Poland	Germany	Luxembourg	United States	Hungary	Spain	Slovak Republic	Latvia	Portugal	Italy	Greece	Turkey	Russian Fed.	Serbia	Thailand	Uruguay	Mexico	Indonesia	Brazil	Tunisia
OECD countries	Upper rank	16	16	17	17	18	18	24		24	25	26	28					29			
	Lower rank	22	23	22	23	23	23	25		26	27	27	28					29			
All countries	Upper rank	19	19	20	20	21	21	27	27	27	29	30	32	32	33	34	35	37	37	38	40
	Lower rank	25	26	25	26	26	26	29	29	30	31	31	33	34	35	36	36	38	39	39	40

* Because data are based on samples, it is not possible to report exact rank order positions for countries. However, it is possible to report the range of rank order positions within which the country mean lies with 95 per cent likelihood.

Instructions:

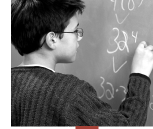
Read across the row for a country to compare performance with the countries listed along the top of the chart. The symbols indicate whether the average performance of the country in the row is lower than that of the comparison country, higher than that of the comparison country, or if there is no statistically significant difference between the average achievement of the two countries.

Without the Bonferroni adjustment:
- Mean performance statistically significantly higher than in comparison country
- No statistically significant difference from comparison country
- Mean performance statistically significantly lower than in comparison country

With the Bonferroni adjustment:
- ▲ Mean performance statistically significantly higher than in comparison country
- ● No statistically significant difference from comparison country
- ▽ Mean performance statistically significantly lower than in comparison country

- Statistically significantly above the OECD average
- Not statistically significantly different from the OECD average
- Statistically significantly below the OECD average

Source: OECD PISA 2003 database.



Consistent with what was found in the other scales, males also show an advantage in the uncertainty scale, particularly at the top end of the distribution (Tables 2.4b and 2.4c). Males outperformed females in 23 OECD countries and six partner countries but differences tend to be small,[18] with an advantage of 11 score points for the combined OECD area. Only in Iceland and the partner country Indonesia did females again outperform males.

Males are slightly ahead of females in the great majority of OECD countries.

It is not possible to compare student performance in 2000 and 2003 on this scale, since the PISA 2000 assessment did not covered this area in its assessment.

OVERALL PERFORMANCE IN MATHEMATICS

The relative strengths and weaknesses of countries in different areas of mathematical content

Comparing performance results in the different content areas of mathematics allows an assessment of the relative strengths and weaknesses of countries. It is not appropriate to compare numerical scale scores directly between the different content areas of mathematics. Nevertheless, it is possible to determine the relative strengths of countries in the different content areas of mathematics on the basis of their relative rank-order positions on the respective scales (Annex A2; Figure A2.1).[19] The values in parenthesis represent mean scores for the space and shape, change and relationships, and the quantity and uncertainty scales, respectively.

In some countries, students show marked differences in their relative performance in different areas of mathematics…

- Student performance on the *space and shape* scale stands out in Japan (553, 536, 527, 528) where it is stronger than on the other three scales, and in Canada (518, 537, 528, 542) and Ireland (476, 506, 502, 517) where the relative standing of these countries is weaker than in the other scales.
- Student performance on the *change and relationships* scale stands out in France (508, 520, 507, 506) while students in the partner countries Hong Kong-China (558, 540, 545, 558) and Macao-China (528, 519, 533, 532) show a lower relative standing on this scale.
- On the *quantity* scale, students in Finland (539, 543, 549, 545) show their strongest performance, while students in New Zealand (525, 526, 511, 532) show their weakest performance on this scale.
- On the *uncertainty* scale, students perform more strongly than on other scales in Greece (437, 436, 446, 458), Iceland (504, 509, 513, 528), Ireland (476, 506, 502, 517), New Zealand (525, 526, 511, 532) and Norway (483, 488, 494, 513). They show a lower relative standing on this scale in Belgium (530, 535, 530, 526), the Czech Republic (527, 515, 528, 500), Germany (500, 507, 514, 493), the Slovak Republic (505, 494, 513, 476) and Switzerland (540, 523, 533, 517) as well as in the partner countries Liechtenstein (538, 540, 534, 523) and the Russian Federation (474, 477, 472, 436).

...and while seven OECD countries have very similar results across content areas, 11 show especially great differences...

The relative standing of some countries, most notably Greece, Italy, Korea, Mexico, Portugal, Spain and Turkey, is very similar across the four mathematics content areas. By contrast, in Austria, Canada, the Czech Republic, France, Germany, Ireland, Japan, New Zealand, Norway, the Slovak Republic and Switzerland, performance differences among the scales are particularly large and may warrant attention in curriculum development and implementation. For example, among OECD countries, the Slovak Republic ranks around fourteenth (twelfth to seventeenth) and thirteenth (ninth to seventeenth) for the space and shape and quantity scales, but around twenty-fourth (twenty-fourth to twenty-fifth) in the uncertainty scale. Similarly, the Czech Republic ranks around seventh (fifth to ninth) on the space and shape scale and around fifth (third to eighth) on the quantity scale but around sixteenth (fifteenth to eighteenth) on the uncertainty scale. New Zealand ranks around sixth (fourth to eighth) and seventh (fifth to ninth) on the uncertainty and space and shape scales, but around sixteenth (eleventh to seventeenth) on the quantity scale. Switzerland ranks third (third to fourth) and fourth (second to seventh) on the space and shape and quantity scales only twelfth (tenth to fourteenth) on the uncertainty scale.

...and in some cases this makes overall mathematics performance seem somewhat lower than in the narrower assessment in 2000.

For some countries – most notably Japan – the relative standing in the content areas that were also assessed in 2000 remained broadly similar while performance was lower on the quantity and uncertainty scales that were newly introduced in 2003. While it would thus be wrong to conclude that mathematics performance in these countries has declined, the results do suggest that the introduction of new content areas in the assessment – quantity and uncertainty (essentially because these are valued and considered important by member countries in the OECD) – sheds a slightly different light on the overall performance of these countries in 2003.

A summary picture of mathematics performance

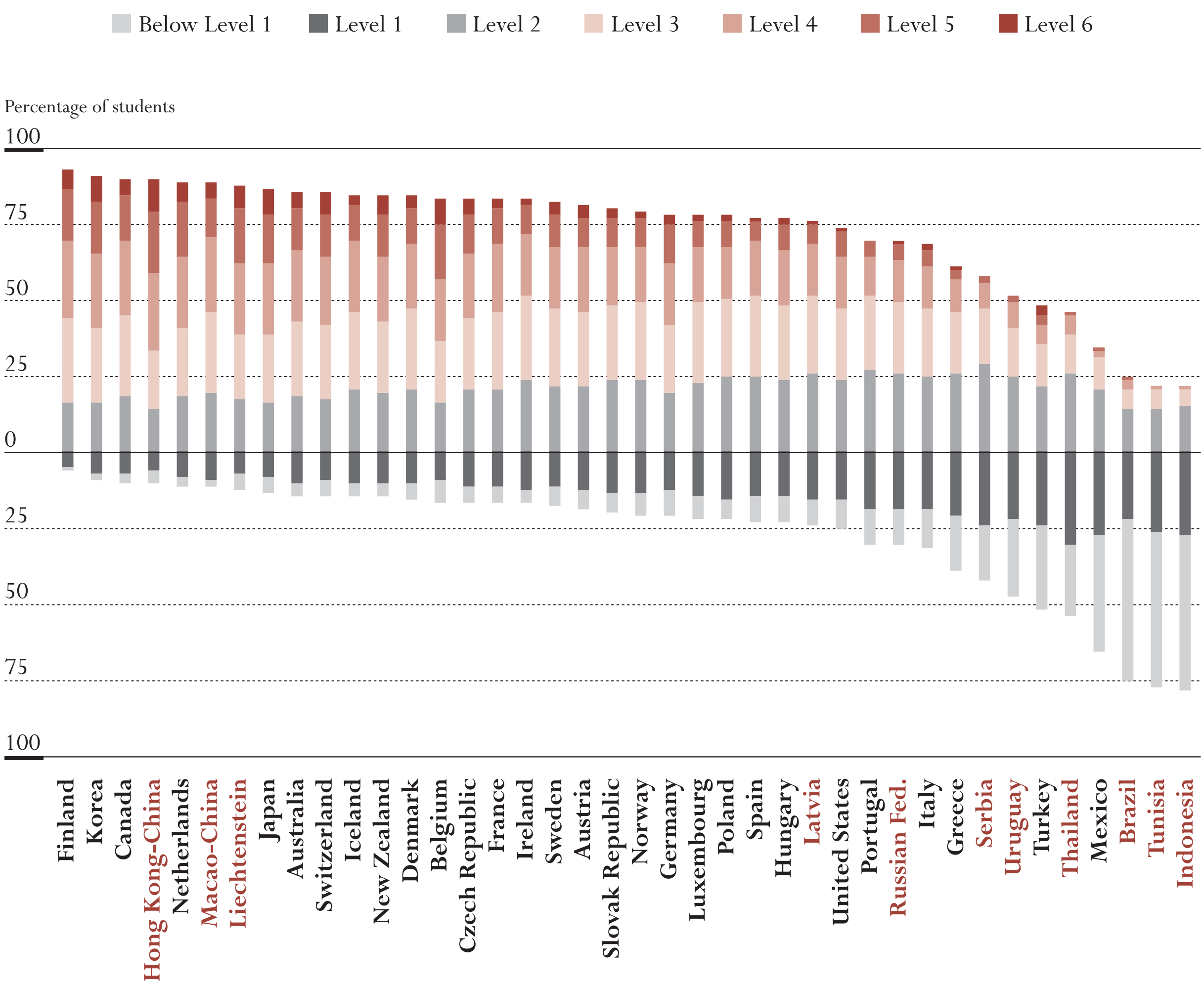
A combined mathematics scale shows performance across the four content areas...

While the relative performance of countries in the four content areas of mathematics is of importance for policy makers as it provides insight into potential strengths and weaknesses of the intended curricula and the effectiveness with which these are delivered, it is also possible to construct a combined performance scale covering performance across the four content areas. Results from this comparison are presented in Figure 2.16a, which shows the percentage of students performing against the international benchmarks defined by the PISA proficiency levels.

...indicating that the top third of students perform at least at Level 4, but the bottom quarter lack all but the basic skills at Level 1...

The results show that about a third of students in OECD countries perform at the top three levels of the mathematics scale (Table 2.5a), but that this figure varies widely in both OECD and the partner countries: half or more of 15-year-olds perform at least at Level 4 in Finland and Korea as well as in the partner country Hong Kong-China. However, only 3 per cent perform at Level 4 in Mexico, with an even lower percentage in the partner countries Indonesia and Tunisia. In most OECD countries, at least three quarters of students perform at or above Level 2. Nevertheless, in Italy, Portugal and the United States over a quarter of students are unable to complete tasks at Level 2. In Greece over a third of students fail

Figure 2.16a ■ **Percentage of students at each level of proficiency on the mathematics scale**



Countries are ranked in descending order of percentage of 15-year-olds in Levels 2, 3, 4, 5 and 6.
Source: OECD PISA 2003 database, Table 2.5a.

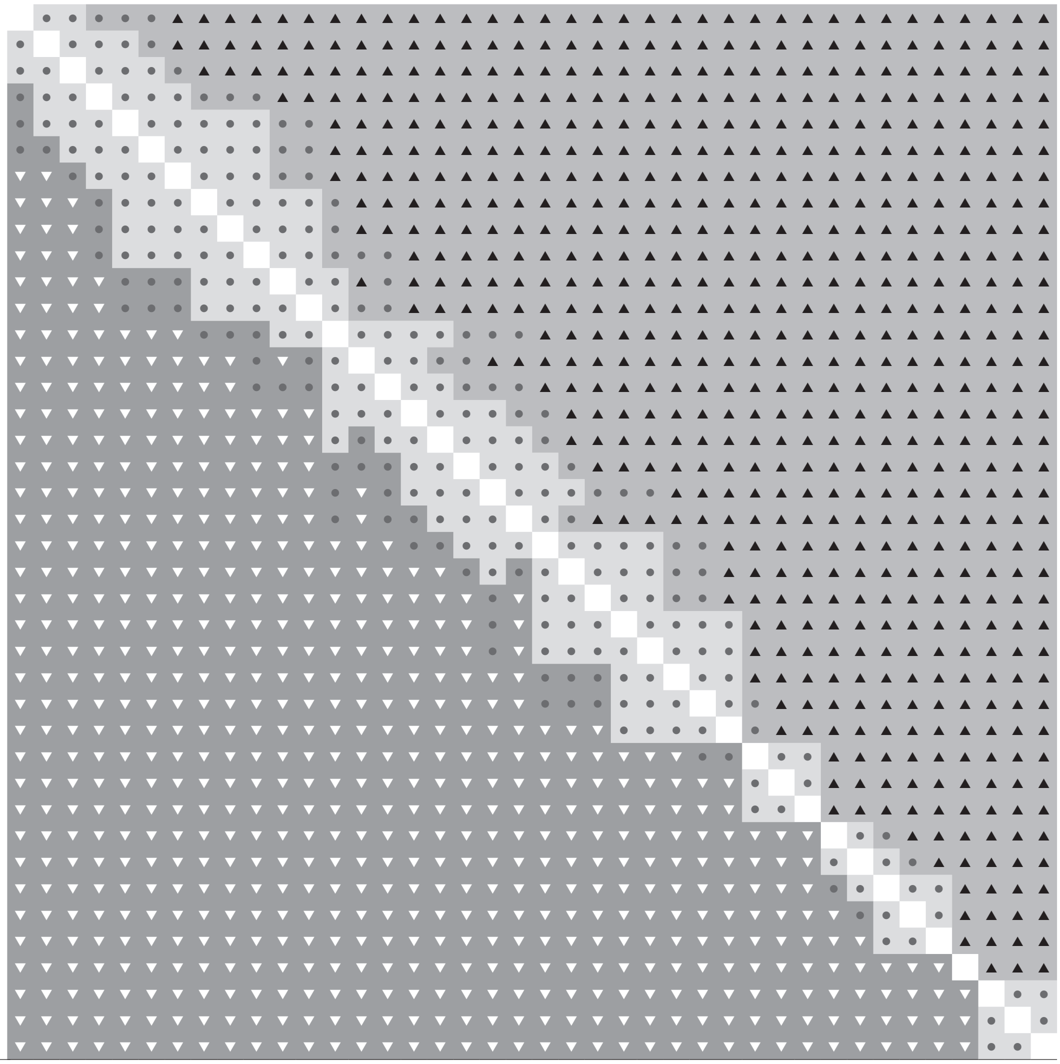
to attain Level 2, and in Mexico and Turkey the majority of students do not achieve this level. These students fail to demonstrate consistently that they have baseline mathematical skills, such as the capacity to use direct inference to recognise the mathematical elements of a situation, use a single representation to help explore and understand a situation, use basic algorithms, formulae and procedures, and the capacity to make literal interpretations and apply direct reasoning (Table 2.5a).

...and these can be combined to compare overall mathematics performance in countries.

Figure 2.16b gives a summary of overall student performance in different countries on the mathematics scale, presented in terms of the mean student score. As discussed in Box 2.1, when interpreting mean performance, only those differences between countries that are statistically significant should be taken into account. The figure therefore shows those pairs of countries where the difference in their mean

Figure 2.16b ■ Multiple comparisons of mean performance on the mathematics scale

Mathematics scale	Mean	S.E.
Hong Kong-China	550	(4.5)
Finland	544	(1.9)
Korea	542	(3.2)
Netherlands	538	(3.1)
Liechtenstein	536	(4.1)
Japan	534	(4.0)
Canada	532	(1.8)
Belgium	529	(2.3)
Macao-China	527	(2.9)
Switzerland	527	(3.4)
Australia	524	(2.1)
New Zealand	523	(2.3)
Czech Republic	516	(3.5)
Iceland	515	(1.4)
Denmark	514	(2.7)
France	511	(2.5)
Sweden	509	(2.6)
Austria	506	(3.3)
Germany	503	(3.3)
Ireland	503	(2.4)
Slovak Republic	498	(3.3)
Norway	495	(2.4)
Luxembourg	493	(1.0)
Poland	490	(2.5)
Hungary	490	(2.8)
Spain	485	(2.4)
Latvia	483	(3.7)
United States	483	(2.9)
Russian Fed.	468	(4.2)
Portugal	466	(3.4)
Italy	466	(3.1)
Greece	445	(3.9)
Serbia	437	(3.8)
Turkey	423	(6.7)
Uruguay	422	(3.3)
Thailand	417	(3.0)
Mexico	385	(3.6)
Indonesia	360	(3.9)
Tunisia	359	(2.5)
Brazil	356	(4.8)



Range of rank*

		Hong Kong-China	Finland	Korea	Netherlands	Liechtenstein	Japan	Canada	Belgium	Macao-China	Switzerland	Australia	New Zealand	Czech Republic	Iceland	Denmark	France	Sweden	Austria	Germany	Ireland	Slovak Republic	Norway	Luxembourg	Poland	Hungary	Spain	Latvia	United States	Russian Fed.	Portugal	Italy	Greece	Serbia	Turkey	Uruguay	Thailand	Mexico	Indonesia	Tunisia	Brazil
OECD countries	Upper rank		1	1	1		2	4	4		4	7	7	9	10	10	11	12	13	14	15	16	18	19	19	19	22		22		25	25	27		28			29			
	Lower rank		3	4	5		7	7	8		9	9	10	14	13	14	15	16	18	18	18	21	21	21	23	23	24		24		26	26	27		28			29			
All countries	Upper rank	1	1	1	2	2	3	5	5	6	6	9	9	12	13	13	14	15	16	17	17	19	21	22	22	22	25	25	25	29	29	29	32	32	33	34	34	37	38	38	38
	Lower rank	3	4	5	7	9	10	9	10	12	12	12	13	17	16	17	18	19	20	21	21	24	24	24	26	27	28	28	28	31	31	31	33	34	36	36	36	37	40	40	40

* Because data are based on samples, it is not possible to report exact rank order positions for countries. However, it is possible to report the range of rank order positions within which the country mean lies with 95 per cent likelihood.

Instructions:

Read across the row for a country to compare performance with the countries listed along the top of the chart. The symbols indicate whether the average performance of the country in the row is lower than that of the comparison country, higher than that of the comparison country, or if there is no statistically significant difference between the average achievement of the two countries.

Without the *Bonferroni adjustment:*
- Mean performance statistically significantly higher than in comparison country
- No statistically significant difference from comparison country
- Mean performance statistically significantly lower than in comparison country

With the *Bonferroni adjustment:*
- ▲ Mean performance statistically significantly higher than in comparison country
- ● No statistically significant difference from comparison country
- ▽ Mean performance statistically significantly lower than in comparison country

- Statistically significantly above the OECD average
- Not statistically significantly different from the OECD average
- Statistically significantly below the OECD average

Source: OECD PISA 2003 database.

scores is sufficient to say with confidence that the higher performance by sampled students in one country holds for the entire population of enrolled 15-year-olds. A country's performance relative to that of the countries listed along the top of the figure can be seen by reading across each row. The colour-coding indicates whether the average performance of the country in the row is either lower than that of the comparison country, not statistically different, or higher. When making multiple comparisons, *e.g.* when comparing the performance of one country with that of all other countries, an even more cautious approach is required, and only those comparisons that are indicated by the upward or downward pointing symbols should be considered statistically significant for the purpose of multiple comparisons. Figure 2.16b also shows which countries perform above, at or below the OECD average.

It is only possible to present a range of ranks for each country…

For the reasons explained in Box 2.1, it is also not possible to determine the exact rank order position of countries in the international comparisons. However, Figure 2.16b shows, with 95 per cent probability, the range of rank order positions within which the country mean lies, both for the group of OECD countries and for all countries that participated in PISA 2003.

…but within-country differences are critical, including some regional differences that can be measured…

Mean performance scores are typically used to assess the quality of schools and education systems. However, it has been noted above that mean performance does not provide a full picture of student performance and can mask significant variation within an individual class, school or education system. The performance variation among schools is examined more closely in Chapter 4. To capture variation between education systems and regions within countries, some countries have also undertaken the PISA assessment at sub-national levels. Where such results are available, these are presented in Annex B2. For some countries, such sub-national differences are very large. For example, mean scores on the mathematics scale for the Flemish community in Belgium are higher than those in the best-performing OECD countries, Finland and Korea. In contrast, the results from the French community are at the OECD average level.

…so it is useful to look at how each country's scores are distributed around their mean…

Figure 2.17 sheds further light on the performance distribution within countries. This analysis needs to be distinguished from the examination of the distribution of student performance across the PISA proficiency levels discussed above. Whereas the distribution of students across proficiency levels indicates the proportion of students in each country that can demonstrate a specified level of knowledge and skills, and thus compares countries on the basis of *absolute* benchmarks of student performance, the analysis below focuses on the *relative* distribution of scores, *i.e.,* the *gap* that exists between students with the highest and the lowest levels of performance *within* each country. This is an important indicator of the equality of educational outcomes in mathematics.

…revealing that each country has students both with very low and very high performance…

The gradation bars in the figure show the range of performance in each country between the 5th percentile (the point below which the lowest-performing 5 per cent of the students in a country score) and the 95th percentile (the point below which 95 per cent of students perform or, alternatively, above which the 5 per cent highest-performing students in a country score). The density of the bar

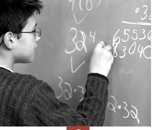
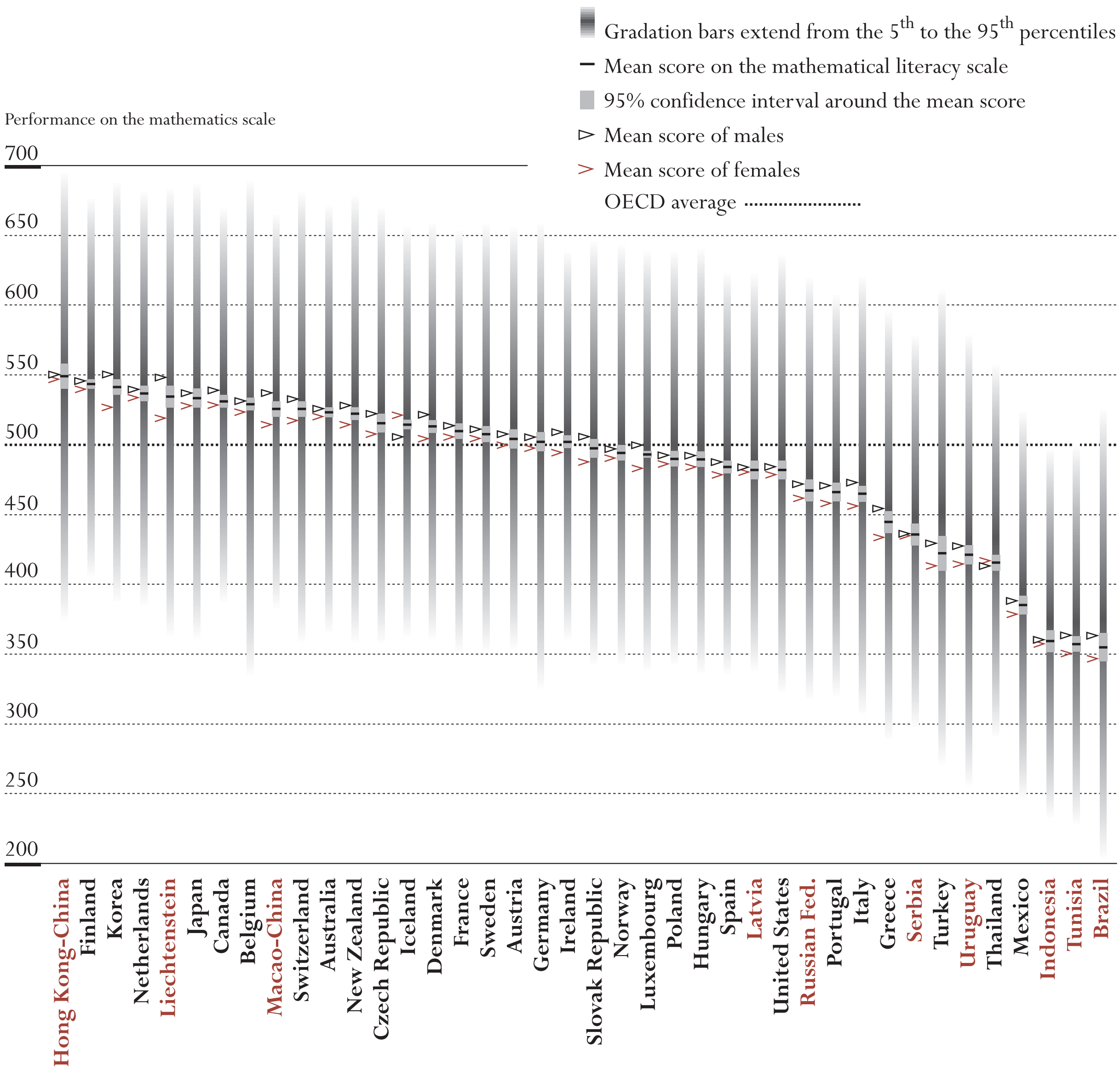
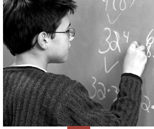


Figure 2.17 ■ **Distribution of student performance on the mathematics scale**



Source: OECD PISA 2003 database, Table 2.5c.

represents the proportion of students performing at the corresponding scale points. The solid, horizontal black line near the middle shows the mean score for each country (*i.e.,* the subject of the discussion in the preceding section) and is located inside a shaded box that shows its confidence interval. The figure shows that there is wide variation in overall student performance on the mathematics scale within countries. The middle 90 per cent of the population shown by the length of the bars exceeds by far the range between the mean scores of the highest and lowest performing countries. In almost all OECD countries, this group includes some students proficient at Level 5 and others not proficient above Level 1. In the majority of countries, the range of performance among the



middle half of the students exceeds the magnitude of two proficiency levels, and in Belgium and Germany it is around 2.4 proficiency levels. This suggests that educational programmes, schools or teachers need to cope with a wide range of student knowledge and skills.

...and that the middle half of students vary in performance...

In addition, Table 2.5c identifies the 25th and 75th percentiles, *i.e.,* the scale points that mark the bottom and top quarters of performers in each country. To what extent are differences in student performance a reflection of a natural distribution of ability and, therefore, difficult to influence through changes in public policy? It is not easy to answer such a question with data from PISA alone, not least because differences between countries are influenced by the social and economic context in which education and learning take place. Nonetheless, several findings suggest that public policy can play a role:

...by more in some countries than others.

- First, the amount of within-country variation in performance in mathematics varies widely between OECD countries. For instance, the difference between the 75th and 25th percentiles ranges from less than 120 score points on the mathematics scale in Canada, Finland, Ireland and Mexico to more than 140 score points in Belgium and Germany. In Belgium, this difference can be explained, at least partially, by the difference in performance between the Flemish and French communities (Annex B2).

Countries with similar levels of average performance show considerable variation in disparities of student performance ...

- Second, countries with similar levels of average performance show a considerable variation in disparity of student performance. For example, Germany and Ireland both score near the OECD average but, while Ireland shows one of the narrowest distributions, the difference between the 75th and 25th percentiles in Germany is among the widest. Similarly, towards the lower end of the scale, Italy and Portugal show similar levels of average performance, but Portugal shows much less performance variation than Italy. And among the top performing countries, Finland displays much less performance variation than Korea or the Netherlands.

...with some high-performing countries managing to limit performance gaps.

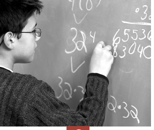
- Third, it is evident from a comparison between the range of performance within a country and its average performance that wide disparities in performance are not a necessary condition for a country to attain a high level of overall performance. As an illustration, Canada, Denmark, Finland, Iceland and Korea all have above-average performance but below-average differences between the 75th and 25th percentiles (Table 2.5c).

Gender differences in mathematics

Previous sections have examined how performance differs among males and females in the different mathematical content areas. This section draws this information together.

Females have made great progress in reducing their historic educational disadvantage, and are ahead in many respects...

Policy-makers have given considerable priority to issues of gender equality, with particular attention being paid to the disadvantages faced by females. Undeniably, significant progress has been achieved in reducing the gender gap in formal educational qualifications. Younger women today are far more likely to have completed a tertiary qualification than women 30 years ago: in 18 of the 29 OECD countries with comparable data, more than twice as many women



aged 25 to 34 have completed tertiary education than women aged 55 to 64 years. Furthermore, university-level graduation rates for women now equal or exceed those for men in 21 of the 27 OECD countries for which data are comparable (OECD, 2004a).

…yet males continue to do better at the tertiary level in mathematics and associated disciplines…

However, in mathematics and computer science, gender differences in tertiary qualifications remain persistently high: the proportion of women among university graduates in mathematics and computer science is only 30 per cent, on average, among OECD countries. In Austria, Belgium, Germany, Hungary, Iceland, the Netherlands, Norway, the Slovak Republic and Switzerland this share is only between 9 and 25 per cent (OECD, 2004a).

…suggesting that schools still have work to do in nurturing performance and interest among females.

Much therefore remains to be done to close the gender gap in mathematics and related fields in tertiary education and evidence suggests that action in this area needs to be targeted at youth and, indeed, children (Box 2.3). At age 15, many students are approaching major transitions from education to work, or to further education. Their performance at school, and their motivation and attitudes towards mathematics, can have a significant influence on their further educational and occupational pathways. These, in turn, can have an impact not only on individual career and salary prospects, but also on the broader effectiveness with which human capital is developed and utilised in OECD economies and societies.

Box 2.3 ■ **Changes in gender differences in mathematics and science performance between lower and upper levels of educational systems**

In 1994-95, the IEA Third International Mathematics and Science Study (TIMSS) revealed statistically significant gender differences in mathematics among fourth-grade students in only three out of the 16 participating OECD countries (Japan, Korea and the Netherlands). In all cases the gender gap favoured males. However, the same study showed statistically significant gender differences in mathematics at the grade-eight level in six of the same 16 OECD countries, all in favour of males. And finally, in the last year of upper secondary schooling, gender differences in mathematics literacy performance in the TIMSS assessment were large and statistically significant in all participating OECD countries, except Hungary and the United States (again, all in favour of males). A similar and even more pronounced picture emerged in science (Beaton *et al.,* 1996; Mullis *et al.,* 1998).

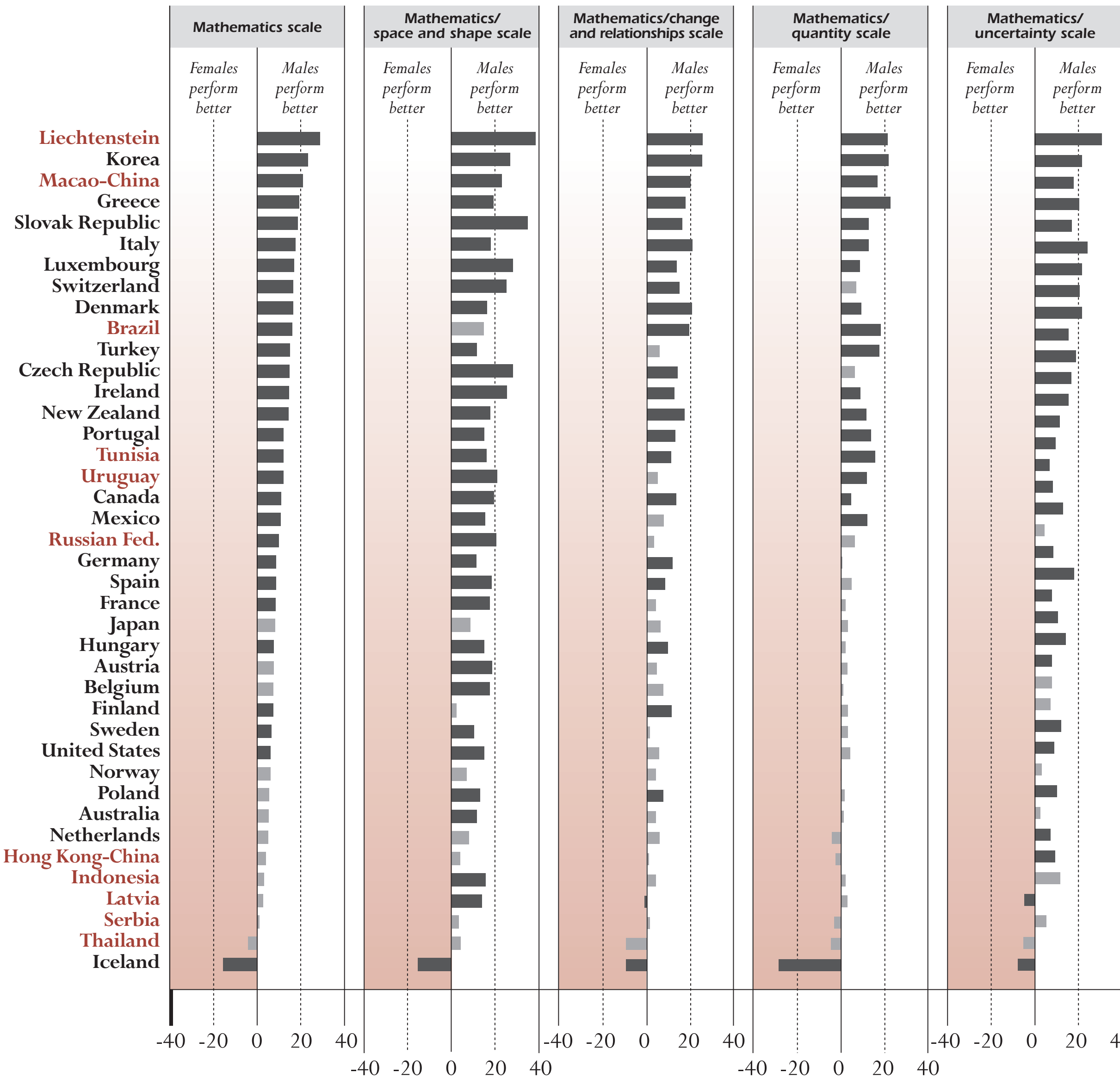
Although the groups of students assessed at the different grade levels were not made up of the same individuals, the results suggest that gender differences in mathematics and science become more pronounced and pervasive in many OECD countries at higher grade levels.

Despite this general tendency, TIMSS also showed that some countries were managing to contain the growth in gender disparities at higher grade levels (OECD, 1996; OECD, 1997).

In this regard, it is striking how closely the broader gender patterns in later career and occupational choices are already mirrored in the mathematics performance of 15-year-old males and females as observed by PISA. And as shown in Chapter 3, gender differences are even more pronounced in the attitudes and approaches towards mathematics shown by 15-year-old males and females. Gender patterns in mathematics performance are fairly consistent across OECD countries (Figure 2.18). Overall, the gender differences appear to be largest in the mathematics/space and shape scale, where performance differences between males and females are visible for all OECD countries except Finland, Norway,

PISA confirms that by age 15, gender differences are visible in most countries, with males performing better, particularly at the high end of the performance distribution.

Figure 2.18 ■ **Gender differences in student performance in mathematics**
Differences in PISA scale scores



Note: Gender differences that are statistically significant are marked in darker colour (see Annex A4).
Source: OECD PISA 2003 database, Tables 2.5c, 2.1c, 2.2c, 2.3c and 2.4c.

the Netherlands and Japan. Gender differences are similarly important in the mathematics/uncertainty scale, where performance differences are visible for 24 out of the 30 OECD countries. Finally, gender differences tend to be larger at the top end of the performance distribution.

While, overall, the gender gap tends to be small,

Iceland is the only OECD country where females consistently perform better than males do. In Australia, Austria, Belgium, Japan, the Netherlands, Norway and Poland, as well as in the partner countries Hong Kong-China, Indonesia, Latvia, Serbia, and Thailand gender differences on the overall mathematics scale are not statistically significant. For the other countries with visible differences, the advantage of males varies widely. In Canada, Denmark, Greece, Ireland, Korea, Luxembourg, New Zealand, Portugal and the Slovak Republic and in the partner countries Liechtenstein, Macao-China and Tunisia, males outperform females in all four content areas, in some of these cases by notable amounts. In contrast, in Austria, Belgium, the United States and the partner country Latvia males outperform females only on the mathematics/space and shape scale, and in Japan, the Netherlands and Norway only on the mathematics/uncertainty scale (Table 2.5c). The percentages of males and females at the lower end of the scale are not consistent across countries. For example, in Iceland, 7 per cent more males than females perform at or below Level 1 while in Greece and Turkey 6 per cent more females than males perform at or below Level 1. On the top end of the scale, in virtually all countries more males than females perform at Level 6 and in the case of Japan and partner country Liechtenstein, this difference is 5 and 7 per cent respectively (Table 2.5b).

Nevertheless, as noted in previous sections, gender differences tend to be small, and are certainly much smaller than the gender differences that were observed by PISA 2000 in the area of reading literacy.[20]

…much larger differences are observed within individual schools…

One issue, however, that needs to be taken into account when interpreting the observed gender differences is that males and females, in many countries at least, make different choices in terms of the schools, tracks and educational programmes they attend. Table 2.5d compares the observed gender difference for all students (column 1) with estimates of gender differences observed within schools (column 2) and estimates of gender differences once various programme and school characteristics have been accounted for. In most countries, the gender differences are larger within schools than they are overall. In Belgium, Germany and Hungary, for example, males have an overall advantage of 8, 9 and 8 score points, respectively, on the mathematics scale, but the average gap increases to 26, 31 and 26 points within schools. In these countries, this is a reflection of the fact that females attend the higher performing, academically oriented tracks and schools at a higher rate than males. If the programme and school characteristics measured by PISA are taken into account,[21] then the estimated gender differences increase even further in many countries (column 3). This leads to an underestimation of the gender differences that are observed within schools. In other words, in these countries more females attend schools and tracks with higher average performance but, within these schools and tracks, they tend to perform lower than males.

From a policy perspective – and for teachers in classrooms – gender differences in mathematics performance, therefore, warrant continued attention. This is the case even if the advantage of males over females within schools and programmes is overshadowed to some extent by the tendency of females to attend higher performing school programmes and tracks.

…with clear implications for teachers…

The significant advantage of males in many countries on at least some of the areas of mathematical content may also be the result of the broader societal and cultural context or of educational policies and practices. Whatever the cause, they suggest that countries are having differing success at eliminating gender gaps, and that males typically remain better at mathematics.

…and perhaps for society more generally.

At the same time, some countries do appear to provide a learning environment that benefits both genders equally, either as a direct result of educational efforts or because of a more favourable societal context or both. The wide variation in gender gaps among countries suggests that the current differences are not the inevitable outcomes of differences between young males and females and that effective policies and practices can overcome what were long taken to be inevitable outcomes of differences between males and females in interests, learning styles and, even, in underlying capacities.

Such differences are not inevitable: some countries avoid them.

THE SOCIO-ECONOMIC CONTEXT OF COUNTRY PERFORMANCE

In as much as it is important to take socio-economic background into account when comparing the performance of any group of students, a comparison of the outcomes of education systems needs to account for countries' economic circumstances and the resources that countries can devote to education. This is done in the following analysis by adjusting the mathematics scale for various social and economic variables at the country level. At the same time such adjustments are always hypothetical and therefore need to be examined with caution. In a global society, the future economic and social prospects of both individuals and countries remains dependent on the results they actually achieve, not on the performance that might result if they were to operate under average social and economic conditions.

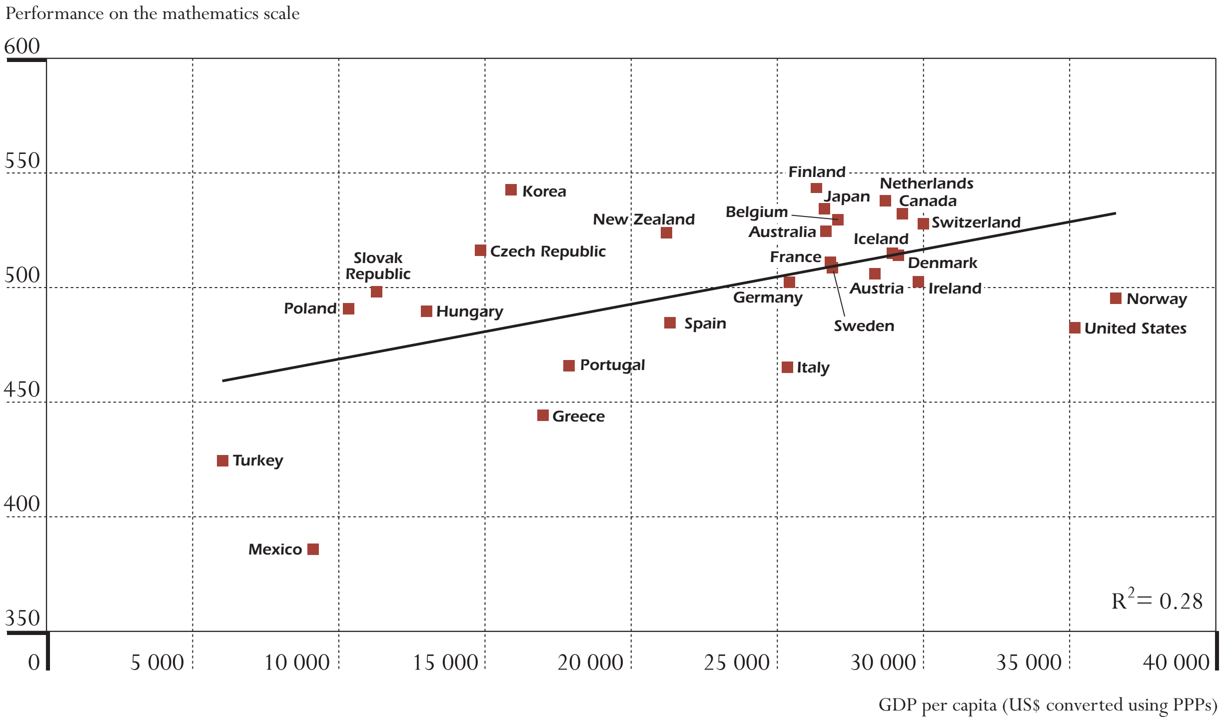
One can also adjust country performance to account for socio-economic differences.

The relative prosperity of some countries allows them to spend more on education, while other countries find themselves constrained by a relatively lower national income. Figure 2.19 displays the relationship between national income as measured by the gross domestic product (GDP) per capita and the average mathematics performance of students in the PISA assessment in each country. The GDP values represent GDP per capita in 2002 at current prices, adjusted for differences in purchasing power between OECD countries (Table 2.6). The figure also shows a trend line that summarises the relationship between GDP per capita and mean student performance in mathematics. It should be borne in mind, however, that the number of countries involved in this comparison is small and that the trend line is therefore strongly affected by the particular characteristics of the countries included in this comparison.

The case for doing so is confirmed by a correlation between national income and mathematics performance, accounting for roughly a fifth of country differences.

Figure 2.19 ■ **Student performance and national income**

Relationship between performance in mathematics and GDP per capita, in US dollars, converted using purchasing power parities (PPPs)



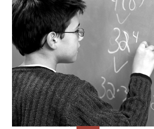
Source: OECD PISA 2003 database, Table 2.6.

The scatter plot suggests that countries with higher national income tend to perform better in mathematics. In fact, the relationship suggests that 28 per cent of the variation between countries' mean scores can be predicted on the basis of their GDP per capita.[22]

There are countries that do better or worse than predicted by their national income.

Countries close to the trend line are where the predictor GDP per capita suggests that they would be. Examples include Austria, Denmark, Germany, Hungary and Sweden. For instance, Sweden outperforms Hungary in mathematics to an extent that one would predict from the difference in their GDP per capita, as shown in Figure 2.19. However, the fact that countries deviate from the trend line also suggests that the relationship is not deterministic and linear. Countries above the trend line have higher average scores on the PISA mathematics assessment than would be predicted on the basis of their GDP per capita (and on the basis of the specific set of countries used for the estimation of the relationship). Countries below the trend line show lower performance than would be predicted from their GDP per capita.

Obviously, the existence of a correlation does not necessarily mean that there is a causal relationship between the two variables; there are, indeed, likely to be many other factors involved. Figure 2.19 does suggest, however, that countries with higher national income are at a relative advantage. This should be taken into



account, in particular, in the interpretation of the performance of countries with comparatively low levels of national income. For some countries, an adjustment for GDP per capita makes a substantial difference to their relative standing internationally. For example, following such an adjustment, Hungary and Poland would move around ten rank order positions upwards on the mathematics scale (490 to 514 and 490 to 521 score points respectively), and the Czech Republic (516 to 536 score points), Portugal (466 to 479 score points) and New Zealand (523 to 528 score points) still by between two and seven positions. Conversely, Austria (506 to 493 score points), Denmark (514 to 500 score points), Norway (495 to 463 score points) and Switzerland (527 to 510 score points) would move between four and six rank positions downwards, given that their performance falls well below what their national levels of income predict.

Adjusting also for adults' educational attainment creates an even greater correction.

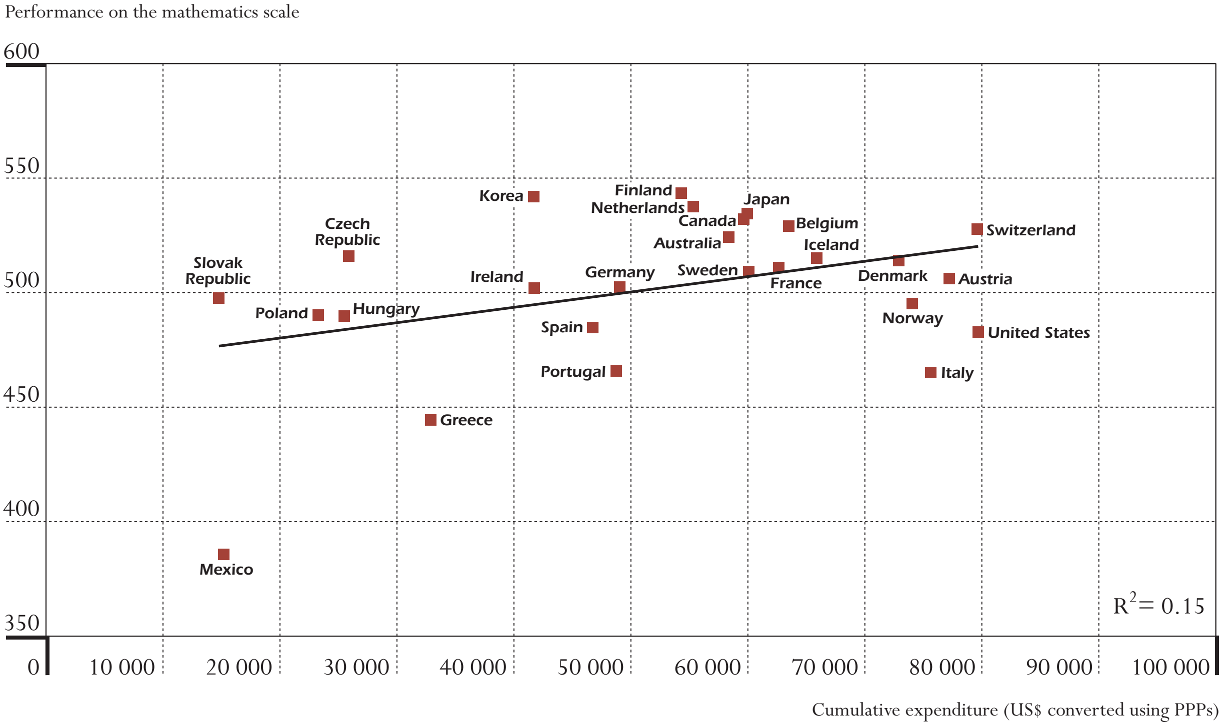
One can further extend the range of contextual variables to be considered further. Given the close interrelationship established in Chapter 4 between student performance and parental levels of educational attainment, an obvious contextual consideration concerns differences in levels of adult educational attainment among the OECD countries. Table 2.6 shows the percentage of the population in the age group 35-44 years that have attained upper secondary and tertiary levels of education. This age group roughly corresponds to the age group of parents of the 15-year-olds assessed in PISA that have attained the upper secondary and tertiary levels of education. If these variables are included in the adjustment in addition to GDP per capita, Poland and Portugal would move upwards by around 16 rank positions (490 to 526 and 466 to 521 score points respectively). Both Poland and Portugal would thus be included in the group of the 10 countries with the highest performance relative to their GDP per capita and levels of adult educational attainment. Conversely, Canada (532 to 510 score points), Denmark (514 to 496 score points), Finland (544 to 525 score points), Germany (503 to 484 score points), Japan (534 to 506 score points), Norway (495 to 459 score points) and Sweden (509 to 487 score points) would move downwards by between 5 and 9 positions, given that their GDP per capita and levels of adult educational attainment would predict far higher levels of student performance than they actually attain. Although combining adult attainment with GDP results in a closer relationship with student performance than when GDP is considered alone, the relationship remains far from deterministic and linear as the model underlying the adjustment assumes. The results therefore need to be interpreted with caution.

Another perspective results from considering how much money countries devote to education…

While GDP per capita reflects the potential resources available for education in each country, it does not directly measure the financial resources actually invested in education. Figure 2.20 compares countries' actual spending per student, on average, from the beginning of primary education up to the age of 15, with average student performance across the three assessment areas. Spending per student is approximated by multiplying public and private expenditure on educational institutions per student in 2002 at each level of education by the theoretical duration of education at the respective level, up to the age of 15.[23] The results are expressed in United States dollars (USD) using purchasing power parities (OECD, 2004a).

Figure 2.20 ■ **Student performance and spending per student**

Relationship between performance in mathematics and cummulative expenditure on educational institutions per student between the ages of 6 and 15 years, in US dollars, converted using purchasing power parities (PPPs)



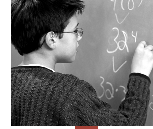
Source: OECD PISA 2003 database, Table 2.6.

...which shows a positive relationship between spending per student and mean mathematics performance...

Figure 2.20 shows a positive relationship between spending per student and mean mathematics performance (see also Table 2.6). As expenditure per student on educational institutions increases, so also does a country's mean performance. However, expenditure per student explains merely 15 per cent of the variation in mean performance between countries.

...but also that high spending levels do not guarantee high performance.

Deviations from the trend line suggest that moderate spending per student cannot automatically be equated with poor performance by education systems. Spending per student between the ages of six and 15 years in the Czech Republic is roughly one-third of, and in Korea roughly one-half of, spending levels in the United States, but while both the Czech Republic and Korea are among the top ten performers in PISA, the United States performs below the OECD average. Similarly, Spain and the United States perform almost equally well, but while the United States spends roughly USD 80 000 per student between the ages of six and 15 years, in Spain this figure is merely USD 47 000. Countries that perform significantly higher than would be expected from their spending per



student alone include Australia, Belgium, Canada, the Czech Republic, Finland, Japan, Korea and the Netherlands. Countries that perform significantly below the level of performance predicted from spending per student include Greece, Italy, Mexico, Norway, Portugal, Spain and the United States. In summary, the results suggest that, while spending on educational institutions is a necessary prerequisite for the provision of high-quality education, spending alone is not sufficient to achieve high levels of outcomes.

IMPLICATIONS FOR POLICY

Mathematics plays a central role for the success of individuals and societies …

For much of the past century, the content of school mathematics curricula was dominated by the need to provide the foundations for the professional training of a small number of mathematicians, scientists and engineers. With the growing role of science, mathematics and technology in modern life, however, the objectives of personal fulfilment, employment and full participation in society increasingly require that all adults, not just those aspiring to a scientific career, should be mathematically, scientifically and technologically literate. The performance of a country's best students in mathematics and related subjects may have implications for the role that the country will play in tomorrow's advanced technology sector, and for its overall international competitiveness. Conversely, deficiencies among lower-performing students in mathematics can have negative consequences for individuals' labour-market and earnings prospects and for their capacity to participate fully in society.

…so most countries attach great importance to securing high performance standards in mathematics throughout their education system…

Not surprisingly, policy-makers and educators alike attach great importance to mathematics education. Addressing the increasing demand for mathematical skills requires excellence throughout education systems, and it is therefore essential to monitor how well countries provide young adults with fundamental skills in this area.

…but some continue to see wide differences in the performance of their students.

The wide disparities in student performance in mathematics within most countries, evident from the analysis in this chapter, suggest that excellence throughout education systems remains still a remote goal and that countries need to serve a wide range of student abilities, including those who perform exceptionally well and also those most in need. At the same time, the analysis has shown that wide disparities in performance are not a necessary condition for a country to attain a high level of overall performance. Indeed, some of the best-performing countries have achieved their results while displaying a modest gap between their stronger and weaker performers.

Relative strengths and weaknesses in various areas of mathematics may lead countries to re-examine curricular priorities.

Performance does not only vary widely among students, but in many countries it also varies between different areas of mathematical content. Such variation may be related to differences in curricular emphases as well as to the effectiveness with which curricula are delivered in different content areas. While countries need to make curricular choices based on their national context and priorities, examining these choices in the light of what other countries consider important can provide a broader frame of reference for national educational policy development.

Gender differences are visible in most countries, with males performing better, particularly at the high end of the performance distribution…

This chapter has shown differences between the performance of males and females in many countries, with the advantage for males being largest in the mathematics/space and shape and the uncertainty scales. Much remains to be done to close the gender gap in mathematics and related fields and evidence suggests that action in this area needs to be targeted at youth and, indeed, children. Their performance at school, and their motivation and attitudes in different subject areas, can have a significant influence on their further educational and occupational pathways. These, in turn, may have an impact not only on individual career and salary prospects, but also on the broader effectiveness with which human capital is developed and utilised in OECD economies and societies. However, the wide variation in gender gaps among countries suggests that the current differences are not the inevitable outcomes of education and that effective policies and practices can overcome what were long taken to be the fixed outcomes of differences in interests, learning styles and even underlying capacities between males and females.

…and, while overall gender differences are often small, the gender gaps which teachers face in classrooms are often considerable.

In most countries, the gender differences are larger within schools than they are overall, reflecting that females tend to attend the higher performing, academically oriented tracks and schools at a higher rate than males but, within these, often perform significantly below males. From a policy perspective – and for teachers in classrooms – gender differences in mathematics performance, therefore, warrant continued attention.

Differences in the overall performance of countries do matter, and cannot be explained only by spending.

Finally, although the variation in student performance within countries is many times larger than the variation between countries, significant differences between countries in the average performance of students should not be overlooked. Particularly in subject areas such as mathematics and science, these differences may raise questions about some countries' future competitiveness. Not all of the variation in the performance of countries in mathematics can be explained by spending on education. Although the analyses have revealed a positive association between the two, they also suggest that while spending on educational institutions is a necessary prerequisite for the provision of high-quality education, spending alone is not sufficient to achieve high levels of outcomes. Other factors, including the effectiveness with which resources are invested, also play a crucial role.

Underperformance matters greatly for individuals, especially where they fail to complete secondary education, reducing their job prospects…

Does mathematics performance on the PISA assessment matter for the future? It is difficult to assess to what extent performance and success in school is predictive of future success. However, what OECD data show is that individuals who have not completed an upper secondary qualification – still roughly one in five on average across OECD countries, despite significant progress over the last generation – face significantly poorer labour-market prospects. For example, labour force participation rates rise steeply with educational attainment in most OECD countries (OECD, 2004a). With very few exceptions, the participation rate for graduates of tertiary education is markedly higher than that for upper secondary graduates which, in turn, is markedly higher than that for individuals without an upper secondary qualification. The gap in male participation rates is particularly wide between upper secondary graduates, and those without an

upper secondary qualification and the labour force participation rate for women with less than upper secondary attainment is particularly low.

...and also their earnings prospects, which tend to be strongly affected by whether they obtain upper secondary and tertiary qualifications...

Similarly, education and earnings are positively linked, with upper secondary education representing a threshold in many countries beyond which additional education attracts a particularly high premium (OECD, 2004a). In all countries, graduates of tertiary-level education earn substantially more than upper secondary graduates. It is possible to contrast, on the one hand, the advantages of education for individuals in terms of higher average earnings, lower risks of unemployment and the public subsidies they receive during their studies with, on the other hand, the costs that individuals incur when studying, in terms of the tuition fees they need to pay, earnings lost during their studies or higher tax rates later in life. The annual rate of return on the investment that individuals incur when completing a tertiary degree is higher than real interest rates, and often significantly so, ranging for males from around 7% in Italy and Japan to 17% in the United Kingdom. Even when public investment in education is included, there is still a positive and significant social return to tertiary education in all countries with comparable data.

...while for society as a whole, education can boost productivity and strengthen citizenship.

In addition, international comparisons show a pivotal role that education plays in fostering labour productivity, and by implication economic growth – not just as an input linking aggregate output to the stock of productive inputs, but also as a factor strongly associated with the rate of technological progress. The estimated long-run effect on economic output of one additional year of education in the combined OECD area is in the order of between 3 and 6 per cent (OECD, 2004a). Finally, the importance of mathematics for citizenship in the modern world should not be overlooked.

Fifteen-year-olds have many chances ahead of them, but those who do well early on are more likely to continue learning...

Obviously, learning does not end with compulsory education and modern societies provide various opportunities for individuals to upgrade their knowledge and skills throughout their lives. However, at least when it comes to job-related continuing education and training, on average across OECD countries, about three times as many training hours are invested in employees with a tertiary qualification, as in employees without an upper secondary qualification (OECD, 2000a and 2000b). Thus, initial education combines with other influences to make job-related training beyond school least likely for those who need it most.

...so poor performance at age 15 causes justifiable concern.

This underlines why a solid foundation of knowledge and skills at school is fundamental for the future success of individuals and societies and the importance of providing opportunities for adults who need to improve their basic levels of literacy in reading, mathematics and science in order to be able to engage in relevant learning throughout their lives. It is in that sense that the results from PISA give rise to concern in many countries.

Notes

1. See Box 2.2 for an explanation.

2. In mathematics, the improvement is statistically significant at the 95 per cent confidence level only for one of the two scales with comparable data.

3. In Mexico, the net enrolment rate of 15-year-olds increased from 51.6 per cent in the 1999-2000 school year to 56.1 per cent in the 2002-03 school year (*Source*: OECD education database).

4. Further technical details on the methods used to estimate student ability and item difficulty, and to form the scale, are provided in the *PISA 2003 Technical Report* (OECD, forthcoming).

5. To be more precise, students were placed at a point on the scale at which they had a 62 per cent chance of answering a question correctly. This is not an arbitrary number: its derivation is related to the definition of proficiency levels, as explained later in this section.

6. Technically, the mean score for student performance in mathematics across OECD countries was set at 500 score points and the standard deviation at 100 score points, with the data weighted so that each OECD country contributed equally. Note that this anchoring of the scale was implemented for the combination of the four scales. The average mean score and standard deviation of the individual mathematics scales can therefore differ from 500 and 100 score points.

7. Results for the combined OECD area are represented in the tables by the ***OECD total.*** The OECD total takes the OECD countries as a single entity, to which each country contributes in proportion to the number of 15-year-olds enrolled in its schools. It illustrates how a country compares with the OECD area as a whole. By contrast, the ***OECD average***, that is also referred to in this report, is the mean of the data values for all OECD countries for which data are available or can be estimated. The OECD average can be used to see how a country compares on a given indicator with a typical OECD country. The OECD average does not take into account the absolute size of the student population in each country, *i.e.,* each country contributes equally to the average. In this publication, the OECD total is generally used when references to the stock of human capital in the OECD area are made. Where the focus is on comparing performance across education systems, the OECD average is used.

8. For the country Serbia and Montenegro, data for Montenegro are not available. The latter accounts for 7.9 per cent of the national population. The name "Serbia" is used as a shorthand for the Serbian part of Serbia and Montenegro.

9. Although the probability that a particular difference will falsely be declared to be statistically significant is low (5 per cent) in each single comparison, the probability of making such an error increases when several comparisons are made simultaneously. It is possible to make an adjustment for this which reduces to 5 per cent the maximum probability that differences will be falsely declared as statistically significant at least once among all the comparisons that are made. Such an adjustment, based on the Bonferroni method, has been incorporated into the multiple comparison charts in this volume, as indicated by the arrow symbols.

10. Column 1 in Table A1.2 estimates the score point difference that is associated with one school year. This difference can be estimated for the 26 OECD countries in which a sizeable number of 15-year-olds in the PISA samples were enrolled in at least two different grades. Since 15-year-olds cannot be assumed to be distributed at random across the grade levels, adjustments had to be made for contextual factors that may relate to the assignment of students to the different grade levels. These adjustments are documented in columns 2 to 7 of the table. While it is possible to estimate the typical performance difference among students in two adjacent grades net of the effects of selection and contextual factors, this difference cannot automatically be equated with the progress that students have made over the last school year but should be interpreted as a lower bound of the progress achieved. This is not only because different students were assessed but also because the contents of the PISA assessment was not expressly designed to match what students had learned in the preceding school year but was designed more broadly to assess the cumulative outcome of learning in school up to age 15. For example, if the curriculum of the grades in which 15-year-olds are enrolled mainly in covers other material than that assessed by PISA (which, in turn, may have been included in earlier school years) then the observed performance difference will underestimate student progress. Accurate measures of student progress can only be obtained through a longitudinal assessment design that focuses on content.

11. When measured in terms of effect sizes (for an explanation of the concept and its interpretation see Box 3.3), these are greater than 0.2 only in Canada, Ireland, Luxembourg, Korea, the Slovak Republic, Spain and Switzerland as well as in the partner countries Liechtenstein, Uruguay and Macao-China. In all countries except Liechtenstein the effect sizes remain below 0.3.

12. See Annex A8 for an explanation of the methods employed to establish the link between the PISA 2000 and 2003 assessments.

13. Luxembourg also shows a significant performance difference. However, the results are not comparable because of changes in assessment conditions. In PISA 2000, students in Luxembourg were given one assessment booklet, with the languages chosen by the students one week prior to the assessment. In practice, however, familiarity with the language of assessment became an important barrier for a significant proportion of students in PISA 2000. In 2003, students were each given one assessment booklet in both languages of instruction and could choose their preferred language immediately prior to the assessment. This provided for assessment conditions that are better comparable with those in countries that have only one language of instruction and results in a fairer assessment of the true performance of students in mathematics, science, reading and problem-solving. As a result of this change in procedures, the assessment conditions and hence the assessment results for Luxembourg cannot be compared between 2000 and 2003. Results for 2000 have therefore been excluded for Luxembourg from this report.

14. In the United States, large standard errors in 2000 may account at least in part for the fact that the United States score is not statistically significantly different between 2000 and 2003.

15. When measured in terms of effect sizes (for an explanation of the concept and its interpretation see Box 3.3), these are greater than 0.2 only in Denmark, Italy and Korea as well as the partner countries Liechtenstein and Macao-China. In all countries the effect sizes remain below 0.3.

16. Also, Luxembourg shows a large performance difference between the 2000 and 2003 results, but – as explained previously – this may be largely due to the modified assessment conditions that allowed students to choose their preferred language from among the two official languages of instruction.

17. When measured in terms of effect sizes (for an explanation of the concept and its interpretation see Box 3.3), these are greater than 0.2 only in Greece, Korea and the partner country Liechtenstein. In all countries the effect sizes remain below 0.3.

18. When measured in terms of effect sizes (for an explanation of the concept and its interpretation see Box 3.3), these are greater than 0.2 only in Denmark, Greece, Korea, Italy, Luxembourg, Switzerland, and the partner countries Liechtenstein and Macao-China. In all countries the effect sizes remain below 0.3.

19. The relative probability of a country assuming each rank-order position on each scale is determined from the country mean scores, their standard errors and the covariance between the performance scales of the two assessment areas. From this it can be concluded whether, with a probability of 95 per cent, a country would rank statistically significantly higher, not statistically differently, or statistically significantly lower on one scale than on the other scale. For details on the methods employed see the *PISA 2003 Technical Report* (OECD, forthcoming).

20. When measured in terms of effect sizes (for an explanation of the concept and its interpretation see Box 3.3), gender differences on the mathematics scale are greater than 0.2 only in Greece, Korea and the partner countries Liechtenstein and Macao-China. In all countries the effect sizes remain below 0.3.

21. A list of the school factors and an explanation of the model used is given in Chapter 5.

22. For the 30 OECD countries included in this comparison, the correlation between mean student performance in mathematics and GDP per capita is 0.43. The explained variation is obtained as the square of the correlation.

23. Cumulative expenditure for a given country is approximated as follows: let $n(0)$, $n(1)$ and $n(2)$ be the typical number of years spent by a student from the age of six up to the age of 15 years in primary, lower secondary and upper secondary education. Let $E(0)$, $E(1)$ and $E(2)$ be the annual expenditure per student in US dollars converted using purchasing power parities in primary, lower secondary and upper secondary education, respectively. The cumulative expenditure is then calculated by multiplying current annual expenditure E by the typical duration of study n for each level of education i using the following formula:

$$CE = \sum_{i=0}^{2} n(i) * E(i)$$

Estimates for $n(i)$ are based on the International Standard Classification of Education (ISCED) (OECD, 1997).