6

A Profile of Student Performance in Reading and Science

Introduction 272

How reading literacy is measured in PISA 272

Student performance in reading 273

- Proficiency at Level 5 (above 625 score points) 276
- Proficiency at Level 4 (from 553 to 625 score points) 278
- Proficiency at Level 3 (from 481 to 552 score points) 278
- Proficiency at Level 2 (from 408 to 480 score points) 278
- Proficiency at Level 1 (from 335 to 407 score points) or below (below 335 score points) 279
- The mean performances of countries in reading 280
- Differences in reading performance between PISA 2000 and PISA 2003 282
- Gender differences in reading literacy 284

How science performance is measured in PISA 286

Student performance in science 293

- The mean performances of countries in science 293
- Differences in science performance between PISA 2000 and PISA 2003 295
- Gender differences in science 296

Implications for policy 298

- Reading 298
- Science 299

INTRODUCTION

The 2003 survey provides an update of reading and science performance.

In PISA 2003, the areas of reading and science were given smaller amounts of assessment time than mathematics (the focus of the 2003 assessment), with 60 minutes for each allowing an update on overall performance rather than the kind of in-depth analysis of knowledge and skills shown for mathematics in Chapter 2. This chapter describes how PISA 2003 measures student achievement in reading and science, examines student outcomes in these two areas, and also compares outcomes for PISA 2003 with PISA 2000.

HOW READING LITERACY IS MEASURED IN PISA

PISA measures students' applied ability to deal with written material…

Reading literacy focuses on the ability of students to use written information in situations which they encounter in their life. In PISA, reading literacy is defined as understanding, using and reflecting written texts, in order to achieve one's goals, to develop one's knowledge and potential and to participate in society. This definition goes beyond the traditional notion of decoding information and literal interpretation of what is written towards more applied tasks.

The concept of reading literacy in PISA is defined by three dimensions: the *format* of the reading material, the *type* of reading task or reading aspects, and the *situation* or the use for which the text was constructed.

…through handling different kinds of texts…

The first dimension, the text format, classifies the reading material or texts into continuous and non-continuous texts. Continuous texts are typically composed of sentences that are, in turn, organised into paragraphs. These may fit under larger structures such as sections, chapters and books. Non-continuous texts are organised differently from continuous texts as they require a different reading approach and can be classified according to their format. Outcomes of students on two reading scales based on the form of the text were reported in the PISA 2000 report *Reading for Change – Performance and Engagement across Countries* (OECD, 2002b).

…and performing different types of reading tasks…

The second dimension is defined by the three reading aspects. Some tasks required students to retrieve information – that is, to locate single or multiple pieces of information in a text. Other tasks required students to interpret texts – that is, to construct meaning and draw inferences from written information. The third type of task required students to reflect on and evaluate texts – that is, to relate written information to their prior knowledge, ideas and experiences. In PISA 2000 student performance in these three types of task were each reported on a separate scale. In 2003, however, less assessment time was allocated to reading and results are reported only on a single reading literacy scale that combines the three types of tasks.

…in relation to various situations where reading is needed.

The third dimension, the situation or context, reflects the categorisation of texts based on the author's intended use, the relationship with other persons implicitly or explicitly associated with the text, and the general content. The situations included in PISA and selected to maximise the diversity of content included in the reading literacy assessment were reading for private use (personal), reading for public use, reading for work (occupational) and reading for education.

A full description of the conceptual framework underlying the PISA assessment of reading literacy is provided in *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills* (OECD, 2003e).

STUDENT PERFORMANCE IN READING

PISA 2003 measures reading in the framework established in 2000, using a subset of tasks used in the PISA 2000 assessment…

The principles for the reporting of results in reading are similar to those applied for mathematics (see Chapter 2). However, unlike in mathematics, where the scales were newly established for the 2003 assessment, the PISA 2003 reading scale is anchored to the results of the 2000 assessment. Since reading was the focus of the 2000 assessment, it was possible to fully develop the instrument for measuring reading literacy at that stage, so the PISA 2000 mean of 500 has been established as the benchmark against which future reading performance will be measured. For reading literacy, PISA 2003 uses an identical framework and a subset of items from PISA 2000. To ensure comparability in calculating trends, the 28 reading items used in PISA 2003 are a subset of the 141 items used in 2000. The subset of items was selected taking the relative balance of aspects of the framework into account; for example, in both years, the proportion of items falling into each task classification is similar (see Table A6.2 for the breakdown of items by the various aspects of the framework).

…and reports results on the same scale that was used in 2000.

Therefore, the reading literacy results that are presented in this chapter are based on the reading literacy proficiency scale that was developed for PISA 2000 which had a mean of 500 and a standard deviation of 100 for the 27 OECD countries that participated. The PISA 2003 results include 29 OECD countries – the Slovak Republic and Turkey joined PISA in 2003 and the Netherlands met all technical standards in 2003, while the United Kingdom has been excluded from the results as it failed to reach the technical standards required by PISA 2003. For the 25 OECD countries for which comparable data are available for both the PISA 2000 and 2003 assessments, the average performance has essentially remained unchanged.[1] However, mainly because of the inclusion of new countries in 2003, the overall OECD mean for reading literacy is now 494 score points and the standard deviation is 100 score points.

The scale divides students into five levels of proficiency…

As in 2000, reading scores in 2003 are reported according to five levels of proficiency, corresponding to tasks of varying difficulty. Proficiency levels are defined by tasks sharing common characteristics including conceptual or substantive as well as statistical ones so that tasks within each level meet certain technical specifications (see Chapter 2). Level 5 corresponds to a score of more than 625, Level 4 to scores in the range 553 to 625, Level 3 to scores from 481 to 552, Level 2 to scores from 408 to 480, and Level 1 to scores from 335 to 407.

…according to the difficulty of tasks that they can usually answer correctly…

Students at a particular level not only demonstrate the knowledge and skills associated with that level but also the proficiencies required at lower levels. For example, all students proficient at Level 3 are also proficient at Levels 1 and 2. All students at a given level are expected to answer at least half of the items at that level correctly (see Chapter 2).

Figure 6.1 ■ **Summary descriptions for the five levels of proficiency in reading literacy**

Level	Retrieving information	Interpreting	Reflecting and evaluating
5	Locate and possibly sequence or combine multiple pieces of deeply embedded information, some of which may be outside the main body of the text. Infer which information in the text is relevant to the task. Deal with highly plausible and/or extensive competing information.	Either construe the meaning of nuanced language or demonstrate a full and detailed understanding of a text.	Critically evaluate or hypothesise, drawing on specialised knowledge. Deal with concepts that are contrary to expectations and draw on a deep understanding of long or complex texts.
	Continuous texts: Analyse texts whose discourse structure is not obvious or clearly marked, in order to discern the relationship of specific parts of the text to its implicit theme or intention. ***Non-continuous texts***: Identify patterns among many pieces of information presented in a display which may be long and detailed, sometimes by referring to information external to the display. The reader may need to realise independently that a full understanding of the section of text requires reference to a separate part of the same document, such as a footnote.		
4	Locate and possibly sequence or combine multiple pieces of embedded information, each of which may need to meet multiple criteria, in a text with familiar context or form. Infer which information in the text is relevant to the task.	Use a high level of text-based inference to understand and apply categories in an unfamiliar context, and to construe the meaning of a section of text by taking into account the text as a whole. Deal with ambiguities, ideas that are contrary to expectation and ideas that are negatively worded.	Use formal or public knowledge to hypothesise about or critically evaluate a text. Show accurate understanding of long or complex texts.
	Continuous texts: Follow linguistic or thematic links over several paragraphs, often in the absence of clear discourse markers, in order to locate, interpret or evaluate embedded information or to infer psychological or metaphysical meaning. ***Non-continuous texts***: Scan a long, detailed text in order to find relevant information, often with little or no assistance from organisers such as labels or special formatting, to locate several pieces of information to be compared or combined.		
3	Locate, and in some cases recognise the relationship between pieces of information, each of which may need to meet multiple criteria. Deal with prominent competing information.	Integrate several parts of a text in order to identify a main idea, understand a relationship or construe the meaning of a word or phrase. Compare, contrast or categorise taking many criteria into account. Deal with competing information.	Make connections or comparisons, give explanations, or evaluate a feature of text. Demonstrate a detailed understanding of the text in relation to familiar, everyday knowledge, or draw on less common knowledge.
	Continuous texts: Use conventions of text organisation, where present, and follow implicit or explicit logical links such as cause and effect relationships across sentences or paragraphs in order to locate, interpret or evaluate information. ***Non-continuous texts***: Consider one display in the light of a second, separate document or display, possibly in a different format, or combine several pieces of spatial, verbal and numeric information in a graph or map to draw conclusions about the information represented.		

Figure 6.1 *(continued)* ■ **Summary descriptions for the five levels of proficiency in reading literacy**

Level	Retrieving information	Interpreting	Reflecting and evaluating
2	Locate one or more pieces of information, each of which may be required to meet multiple criteria. Deal with competing information.	Identify the main idea in a text, understand relationships, form or apply simple categories, or construe meaning within a limited part of the text when the information is not prominent and low-level inferences are required.	Make a comparison or connections between the text and outside knowledge, or explain a feature of the text by drawing on personal experience and attitudes.
	Continuous texts: Follow logical and linguistic connections within a paragraph in order to locate or interpret information; or synthesise information across texts or parts of a text in order to infer the author's purpose. ***Non-continuous texts***: Demonstrate a grasp of the underlying structure of a visual display such as a simple tree diagram or table, or combine two pieces of information from a graph or table.		
1	Locate one or more independent pieces of explicitly stated information, typically meeting a single criterion, with little or no competing information in the text.	Recognise the main theme or author's purpose in a text about a familiar topic, when the required information in the text is prominent.	Make a simple connection between information in the text and common, everyday knowledge.
	Continuous texts: Use redundancy, paragraph headings or common print conventions to form an impression of the main idea of the text, or to locate information stated explicitly within a short section of text. ***Non-continuous texts***: Focus on discrete pieces of information, usually within a single display such as a simple map, a line graph or a bar graph that presents only a small amount of information in a straightforward way, and in which most of the verbal text is limited to a small number of words or phrases.		

...plus a sixth group made up of those unable to show basic functional reading skills.

Students scoring below 335 score points, *i.e.,* those who do not reach Level 1, are not able to routinely show the most basic skills that PISA seeks to measure. While such performance should not be interpreted to mean that those students have no literacy skills at all, performance below Level 1 does signal serious deficiencies in students' ability to use reading literacy as a tool for the acquisition of knowledge and skills in other areas. Similarly, since Level 5 is also unbounded, some students participating in PISA may demonstrate higher reading skills than those measured by the assessment.

Tasks in each proficiency level have identifiable features...

The establishment of proficiency levels in reading makes it possible not only to rank students' performance but also to describe what they can do (Figure 6.1). Each successive reading level is associated with tasks of ascending difficulty. The tasks at each level of reading literacy were judged by panels of experts to share certain features and requirements and to differ consistently from tasks at either higher or lower levels. The assumed difficulty of tasks was then validated empirically on the basis of student performance in participating countries.

The reading literacy tasks used in PISA 2003 include the three dimensions previously described and have a diverse range of difficulty. Samples of the reading tasks (a total of 45 items) were released after PISA 2000 and can be found in the publication *Sample Tasks from the PISA 2000 Assessment – Reading, Mathematical and Scientific Literacy* (OECD, 2002c). Each item includes an indication of the dimension being assessed, and a description of the knowledge and skills being assessed. These descriptions provide some insight into the range of processes required of students and the proficiencies which they need to demonstrate to reach different reading levels. Further sample tasks can also be found at *www.pisa.oecd.org*.

…with easier tasks requiring basic handling of simple texts…

Even a cursory review of these items reveals that tasks at the lower end of the scale require very different skills from those at the higher end. A more careful analysis of the range of tasks provides some indication of an ordered set of knowledge-construction skills and strategies. For example, the easiest of these tasks require students to locate explicitly stated information according to a single criterion where there is little, if any, competing information in the text, or to identify the main theme of a familiar text, or make a simple connection between a piece of the text and everyday life. In general, the information is prominent in the text and the text itself is less dense and less complex in structure.

…and harder ones involving increasing complexity and less explicit information.

In contrast, harder retrieval tasks require students to locate and sequence multiple pieces of deeply embedded information, sometimes in accordance with multiple criteria. Often there is competing information in the text that shares some features with the information required for the answer. Similarly, with tasks requiring interpretation or reflection and evaluation, those at the lower end differ from those at the higher end in terms of the process needed to answer them correctly, the degree to which the reading strategies required for a correct answer are signalled in the question or the instructions, the level of complexity and familiarity of the text and the quantity of competing or distracting information present in the text.

Figure 6.2 presents an overall profile of proficiency on the reading literacy scale, with the length of the bars showing the percentage of students proficient at each level.

Proficiency at Level 5 (above 625 score points)

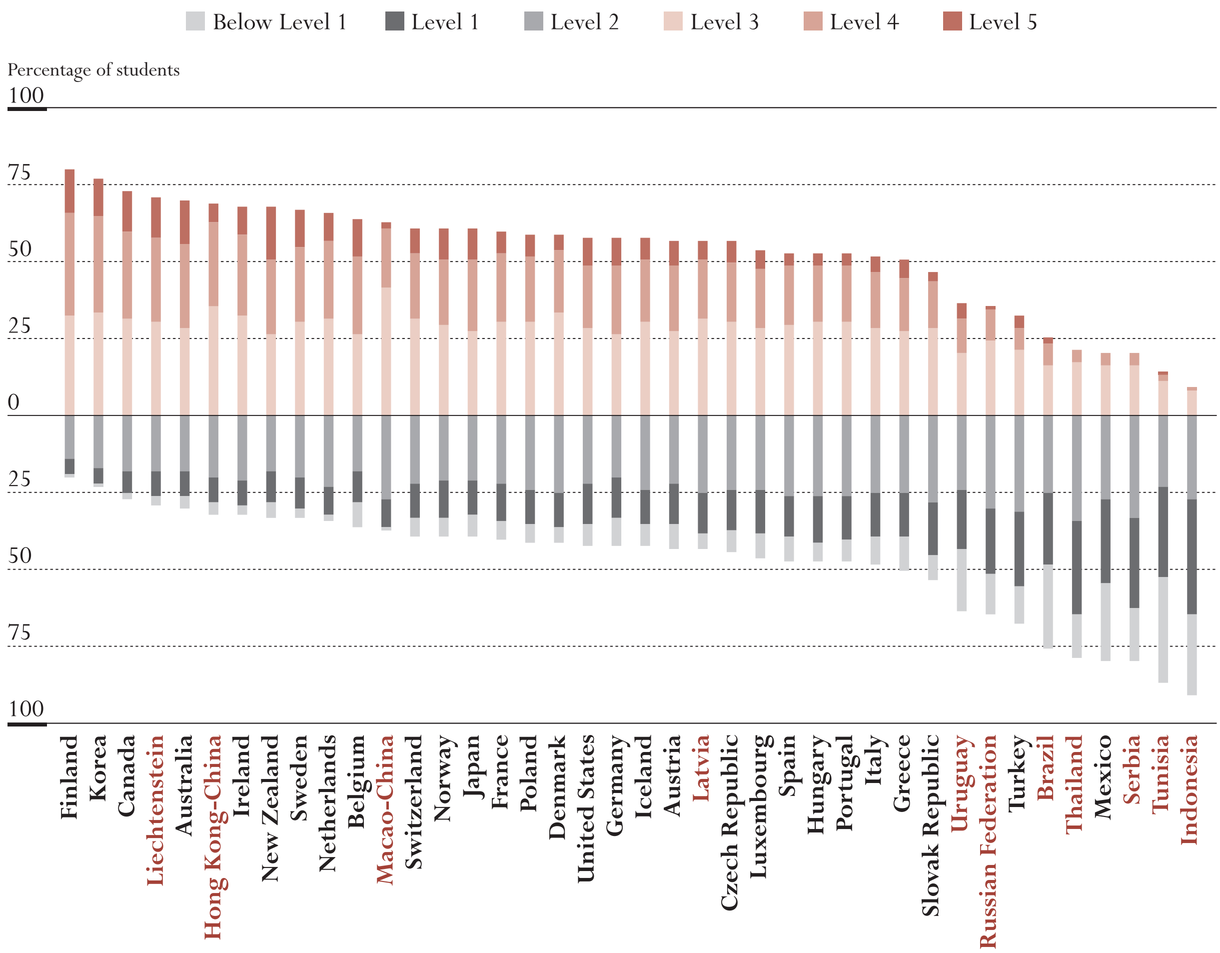
The hardest tasks are sophisticated and require critical thinking…

Students proficient at Level 5 on the reading literacy scale are capable of completing sophisticated reading tasks, such as managing information that is difficult to find in unfamiliar texts; showing detailed understanding of such texts and inferring which information in the text is relevant to the task; and being able to evaluate critically and build hypotheses, draw on specialised knowledge, and accommodate concepts that may be contrary to expectations. See Figure 6.1 for a more detailed description.

…measuring the kind of skill needed by high-level knowledge workers.

The proportion of students performing at the highest PISA proficiency levels in participating countries are of interest as today's proportion of students performing at these levels may influence the contribution which that country

Figure 6.2 ■ **Percentage of students at each level of proficiency on the reading scale**



Countries are ranked in descending order of percentage of 15-year-olds in Levels 3, 4 and 5.
Source: OECD PISA 2003 database, Table 6.1.

will make towards the pool of tomorrow's world-class knowledge workers in the global economy.

In the combined OECD area, 8 per cent of the students are at proficiency Level 5. More than 16 per cent of the students in New Zealand and more than 12 per cent of the students in Australia, Belgium, Canada, Finland, Korea and the partner country Liechtenstein are at this level. In contrast, less than 1 per cent of the students in Mexico reach Level 5 and this is also true in the partner countries Indonesia, Serbia,[2] Thailand and Tunisia (Figure 6.2 and Table 6.1).

Having more students at this top level does not always go with having fewer at lower performance levels.

It is important to keep in mind that the proportion of students performing at Level 5 is influenced not only by the overall performance of countries in reading literacy but also by the variation that exists within countries between the students with the highest and the lowest levels of performance. While there is a general tendency for countries with a higher proportion of students scoring

at Level 5 to have fewer students at Level 1 and below, this is not always the case. In Finland, for example, 15 per cent of students reach Level 5 while only 1 per cent are below Level 1. By contrast, in Belgium and New Zealand, which also have high percentages reaching Level 5, a relatively high proportion of students score below Level 1 as well (8 and 5 per cent respectively). Finally, in the partner countries Hong Kong-China and Macao-China, only 6 per cent and 2 per cent, respectively, reach Level 5, while only 3 per cent and 1 per cent, respectively, score below Level 1.

Proficiency at Level 4 (from 553 to 625 score points)

In some countries around 40 per cent of students can at least do difficult tasks at Level 4, but in others very few can.

Students proficient at Level 4 on the reading literacy scale are capable of difficult reading tasks, such as locating embedded information, dealing with ambiguities and critically evaluating a text (Figure 6.1). In the combined OECD area, 28 per cent of students are proficient at Level 4 or above (that is, at Levels 4 and 5) (Figure 6.2 and Table 6.1). Nearly half of the students in Finland and between 40 and 50 per cent or more of those in Australia, Canada, Korea and New Zealand and the partner country Liechtenstein attain at least Level 4. With the exception of Mexico, the Slovak Republic and Turkey, at least one in five students in each OECD country reaches at least Level 4. In addition, fewer than 5 per cent of the students in four of the partner countries – Indonesia, Serbia, Thailand and Tunisia – reach this level.

Proficiency at Level 3 (from 481 to 552 score points)

Most students in OECD countries have at least moderate reading skills…

Students proficient at Level 3 on the reading literacy scale are capable of reading tasks of moderate complexity, such as locating multiple pieces of information, making links between different parts of a text and relating it to familiar everyday knowledge (Figure 6.1). In the combined OECD area, 55 per cent of students are proficient at least at Level 3 (that is, at Levels 3, 4 and 5) on the reading literacy scale (Figure 6.2 and Table 6.1). In 8 of the 30 OECD countries (Australia, Canada, Finland, Ireland, Korea, the Netherlands, New Zealand and Sweden), and in two partner countries (Hong Kong-China and Liechtenstein), between 65 and 80 per cent of 15-year-old students are proficient at least at Level 3. This level is the OECD modal level – that is, the one at which most students are placed at their highest level of proficiency, with 27 per cent in the OECD combined area.

Proficiency at Level 2 (from 408 to 480 score points)

…and in all but two OECD countries, at least 75 per cent can do basic reading tasks.

Students proficient at Level 2 are capable of basic reading tasks, such as locating straightforward information, making low-level inferences of various types, working out what a well-defined part of a text means and using some outside knowledge to understand it (Figure 6.1). In the combined OECD area, 78 per cent of students are proficient at Level 2 or above on the reading literacy scale. In every OECD country except Mexico and Turkey, at least three in four students are at Level 2 or above (Figure 6.2 and Table 6.1).

Proficiency at Level 1 (from 335 to 407 score points) or below (below 335 score points)

Level 1 represents the simplest functional reading tasks…

Reading literacy, as defined in PISA, focuses on the knowledge and skills required to apply reading for learning rather than on the technical skills acquired in learning to read. Since comparatively few young adults in OECD countries have not acquired technical reading skills, PISA does not seek to measure such things as the extent to which 15-year-old students are fluent readers or how well they spell or recognise words. In line with most contemporary views about reading literacy, PISA focuses on measuring the extent to which individuals are able to construct, expand and reflect on the meaning of what they have read in a wide range of texts common both within and beyond school. The simplest reading tasks that can still be associated with this notion of reading literacy are those at Level 1. Students proficient at this level are capable of completing only the simplest reading tasks developed for PISA, such as locating a single piece of information, identifying the main theme of a text or making a simple connection with everyday knowledge (Figure 6.1).

…and those not reaching it may be able to read but have serious problems using reading for learning.

Students performing below 335 score points – that is, below Level 1 – are not likely to demonstrate success on the most basic type of reading that PISA seeks to measure. This does not mean that they have no literacy skills. Nonetheless, their pattern of answers in the assessment is such that they would be expected to solve fewer than half of the tasks in a test made up of items drawn solely from Level 1, and therefore perform below Level 1. Such students have serious difficulties in using reading literacy as an effective tool to advance and extend their knowledge and skills in other areas. Students with literacy skills below Level 1 may therefore be at risk not only of difficulties in their initial transition from education to work, but also of failure to benefit from further education and learning opportunities throughout life.

Although over nine out of ten OECD students can at least perform at Level 1…

In the combined OECD area, 14 per cent of students perform at Level 1, and 8 per cent perform below Level 1, but there are wide differences between countries. In Finland and Korea, only 5 per cent of students perform at Level 1, and 1 per cent below it, but these countries are the exceptions. In all other OECD countries, the percentage of students performing at or below Level 1 ranges from 10 to 52 per cent (Figure 6.2 and Table 6.1). One-quarter of the OECD countries have between 2 and 5 per cent of students performing below Level 1.

…in 11 OECD countries at least one in five are not proficient beyond Level 1.

The OECD countries with 20 per cent or more of students at or below Level 1 are (in descending order): Mexico, Turkey, Greece, the Slovak Republic, Italy, Luxembourg, Germany, Portugal, Spain, Austria and Hungary. This is also the case in the following partner countries (in descending order): Indonesia, Tunisia, Brazil, Serbia, Thailand, Uruguay and the Russian Federation. It is notable that among these countries Germany has the relatively high contrasting figure of close to 10 per cent of its students performing at Level 5.

In addition, between 25 and 34 per cent of students do not reach Level 1 in Mexico and in the partner countries Brazil, Indonesia and Tunisia. These students are routinely unable to show the most basic skills that PISA seeks to measure.

The mean performances of countries in reading

Country performance can be summarised by a mean score…

The discussion above has focused on comparisons of the distribution of student performance between countries. Another way to summarise student performance and to compare the relative standing of countries in reading literacy is by way of their mean scores. Given that high average performance at age 15 is predictive of a highly skilled future workforce, countries with high average performance will have a considerable economic and social advantage.

…but a comparison of country means is only possible where there is a statistically significant difference.

As discussed in Chapter 2, when interpreting mean performance, only those differences between countries which are statistically significant should be taken into account. Figure 6.3 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. Read across the row for a country to compare its performance with the countries listed along the top of the figure. The colour-coding indicates whether the average performance of the country in the row is significantly lower than that of the comparison country, not statistically different, or significantly higher.

When making multiple comparisons – for example, 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 respective symbols in dark shadings should be considered statistically significant for the purpose of multiple comparisons. The figure also shows which countries perform above, at or below the OECD average.

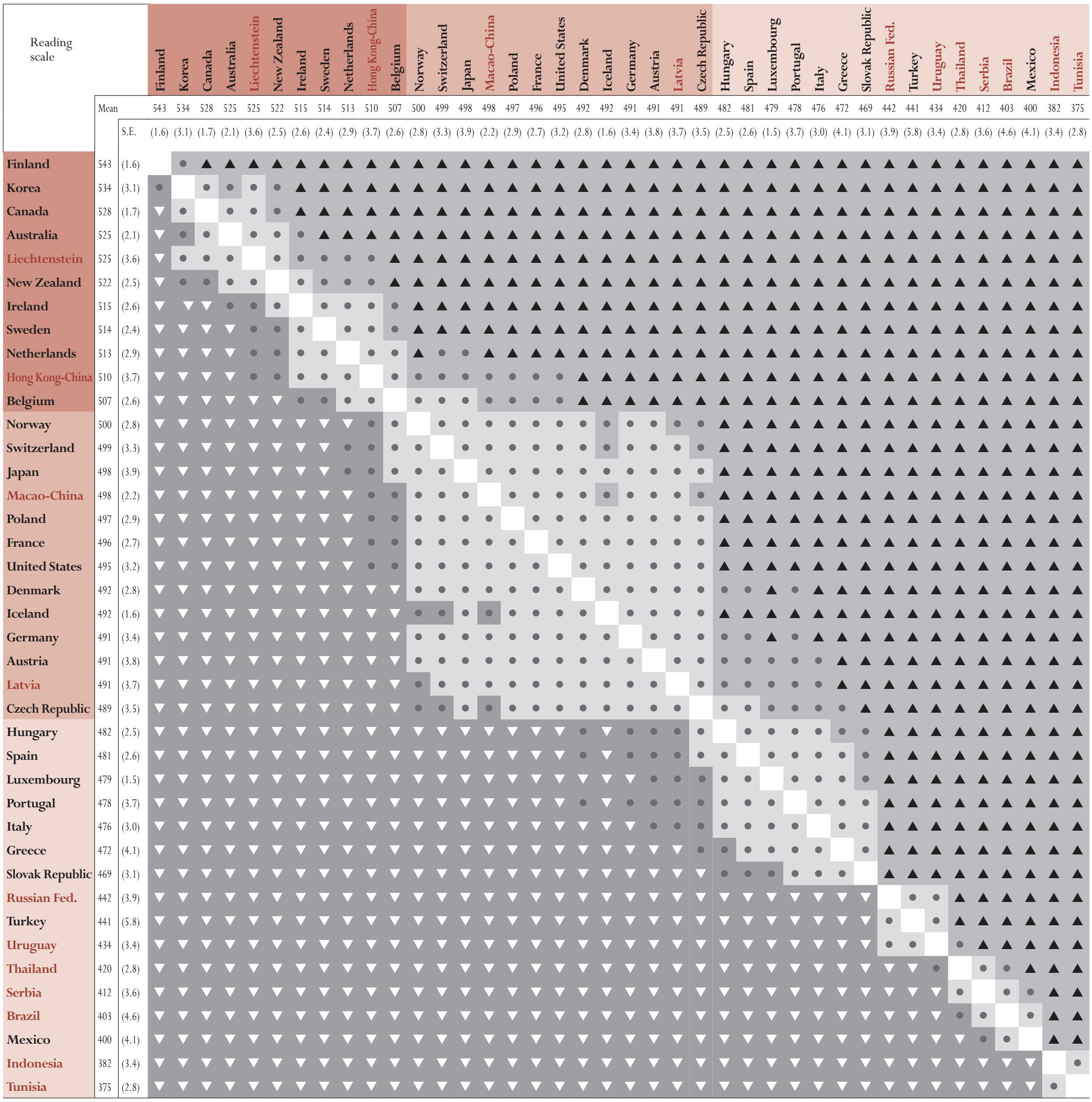
These mean performances span a wide range, with Finnish students doing best overall.

In Finland, performance on the reading literacy scale is above that of any other OECD country. Its country mean, 543 score points, is more than half of a proficiency level above the OECD average of 494 score points in PISA 2003. Other OECD countries with mean performances statistically significantly above the OECD average include Australia, Belgium, Canada, Ireland, Korea, the Netherlands, New Zealand and Sweden. Among the partner countries, Hong Kong-China and Liechtenstein are also part of that group. Eleven OECD countries perform around the OECD average: Austria, the Czech Republic, Denmark, France, Germany, Iceland, Japan, Norway, Poland, Switzerland and the United States. The partner countries Latvia and Macao-China also perform around the OECD average.[3] Among OECD countries, differences are relatively large – 143 score points separate the two extreme performances (*i.e.,* highest and lowest performing countries) – and when the partner countries are considered, this is 150 points.

Within each country, however, the range of performance is even greater, and some countries manage to contain this difference better than others.

Although there are large differences in the mean performance between countries, the variation in performance between students within each country is much larger. One of the major challenges faced by education systems is to encourage high performance while at the same time minimising poor performance. The question of poor performance is particularly relevant to reading literacy because levels of literacy have a significant impact on the welfare of individuals, the state of society and the economic standing of countries in the international arena (OECD, 2003c). Inequality in this context can be examined through the performance distribution as

Figure 6.3 ■ Multiple comparisons of mean performance on the reading scale



Range of rank*

		Finland	Korea	Canada	Australia	Liechtenstein	New Zealand	Ireland	Sweden	Netherlands	Hong Kong-China	Belgium	Norway	Switzerland	Japan	Macao-China	Poland	France	United States	Denmark	Iceland	Germany	Austria	Latvia	Czech Republic	Hungary	Spain	Luxembourg	Portugal	Italy	Greece	Slovak Republic	Russian Fed.	Turkey	Uruguay	Thailand	Serbia	Brazil	Mexico	Indonesia	Tunisia
OECD countries	Upper rank	1	2	2	3		4	6	6	6		8	10	10	10		10	10	10	12	14	12	12		14	20	20	21	21	21	23	25		28					29		
	Lower rank	1	3	4	5		6	8	9	9		10	15	17	18		18	18	19	20	20	20	21		21	24	25	25	26	26	27	27		28					29		
All countries	Upper rank	1	2	2	3	2	4	6	7	7	7	9	11	12	12	12	12	12	12	15	17	15	14	14	17	24	24	25	25	26	27	29	32	32	33	35	35	36	37	39	39
	Lower rank	1	3	5	6	6	7	10	10	11	12	12	18	20	22	19	21	22	23	24	24	24	25	25	25	28	29	29	30	31	31	31	34	34	34	36	37	38	38	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.

seen by the gap in performance between the 5th and the 95th percentiles (Table 6.2). Among OECD countries, Finland and Korea show the narrowest distributions in the OECD with this difference equivalent to 267 score points while at the same time these two countries show the strongest overall performance. From the partner countries, Macao-China has a very narrow distribution with only 220 score points separating the bottom 5th to the top 95th percent of students. Furthermore, in Canada, Denmark, Ireland and the Netherlands and in the partner countries Hong Kong-China, Indonesia, Latvia, Liechtenstein, Serbia and Thailand the performance gaps are below 300 score points. On the other hand, Belgium and Germany show the OECD largest gaps in the performance of students in the middle of the distribution at 362 and 357 score points, which is almost one standard deviation more than in Finland and Korea.

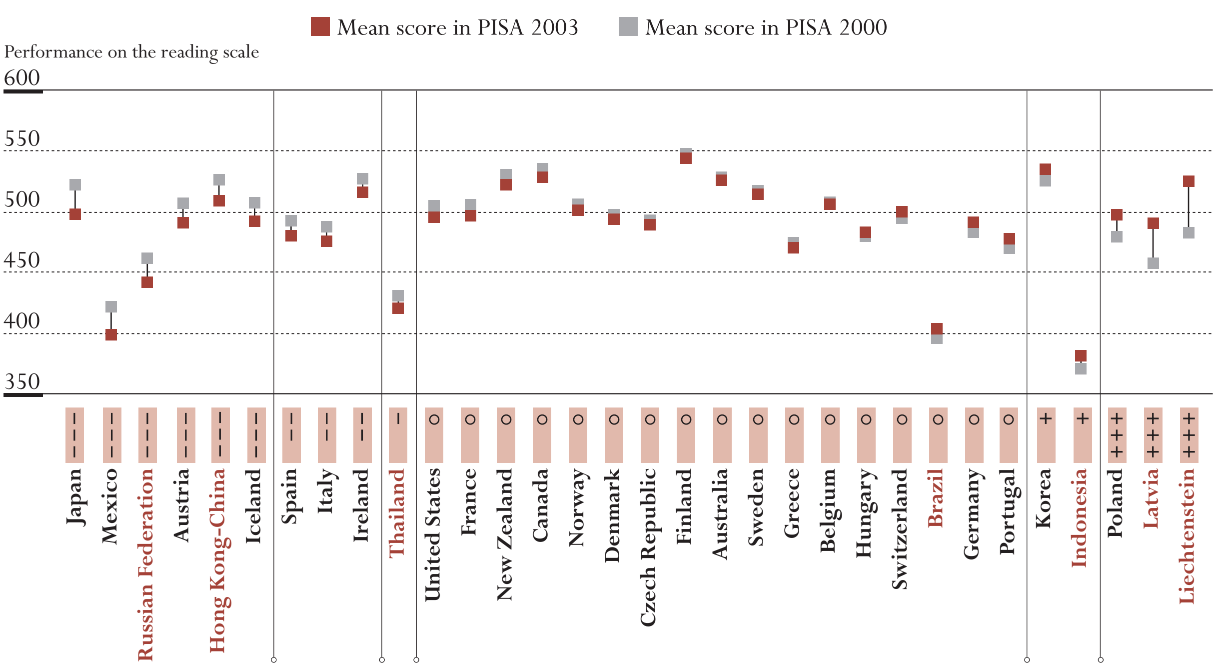
Differences in reading performance between PISA 2000 and PISA 2003

Results from the two PISA surveys should be compared cautiously.

Figure 6.4 shows the overall reading scores for PISA 2000 and 2003 and indicates differences in performance between the two assessments. However, as explained in Chapter 2, such differences need to be interpreted with caution.

Figure 6.4 ■ **Differences in mean scores between PISA 2003 and PISA 2000 on the reading 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 database, Table 6.2; OECD PISA 2000 database, Table 2.3a (OECD 2001a).

First of all, since data are only available from two points in time, it is not possible to assess to what extent the observed differences are indicative of longer-term trends. Furthermore, errors from sampling as well as measurement errors are inevitably introduced when sample-based assessments are linked through a limited number of common items over time, which limits the reliability of comparisons of results over time. To account for the latter, the confidence band for comparisons over time has been broadened correspondingly.[4]

The performance of some countries was slightly better, of others slightly worse.

Figure 6.5 shows that, of the 32 countries for which there is comparative data across 2000 and 2003, in eight there is no statistically significant change at any point in the student distribution. For a further 15 countries, there is a decrease in the scores

Figure 6.5 ■ **Comparisons between PISA 2003 and PISA 2000 in reading**

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

Source: OECD PISA 2003 database, Table 6.2; OECD PISA 2000 database, Table 2.3a (OECD 2001a).

of one or more percentile points, for six countries there is an improvement of one or more points and for only one country the results were mixed.

Poland raised its overall performance through improvements at the lower end of the distribution…

Poland and the partner countries Indonesia, Latvia and Liechtenstein showed markedly higher performance in 2003 than in 2000.[5] In Poland, the overall performance gap between the lower and higher achievers decreased at the same time that the average performance of 15-year-olds increased overall. This rise in overall performance is attributable mainly to an increase in performance at the lower end of the performance distribution (*i.e.,* 5^{th}, 10^{th} and 25^{th} percentiles), in other words, the lowest performing students became better. While in 2000, the lowest 10 per cent of 15-year-olds in Poland scored below 343 score points, in 2003 this changed to 374 score points. The reverse holds for Korea where there was a statistically significant increase in the top half of the distribution between 2000 and 2003 to the extent that only 5 per cent of the students in 2000 reached the performance level that is now reached by the best performing 10 per cent of Korean students. Latvia and Liechtenstein showed increases throughout the distribution.

…whereas in other countries changes at different parts of the performance distribution were insufficient to produce change overall.

Canada, Denmark and Finland showed no measurable overall performance differences between 2000 and 2003. However, in these countries performance at the top end of the distribution (*i.e.,* the 75^{th}, 90^{th} and 95^{th} percentiles) decreased somewhat.

Countries with lower performance in 2003 compared with 2000 include Austria, Iceland, Ireland, Italy, Japan, Mexico and Spain and among the partner countries Hong Kong-China, the Russian Federation and Thailand. For Austria, Iceland, Italy, Japan, and Spain the decline is due to a drop in performance among the 5^{th}, 10^{th} and 25^{th} percentiles (the points under which 5, 10 and 25 per cent of the population score). In other words, in these countries the top end of the distribution performed similarly in 2000 and 2003 but the lower end of the distribution performed markedly lower, making the distribution wider. The Russian Federation is the only country which showed a universal decrease in performance.

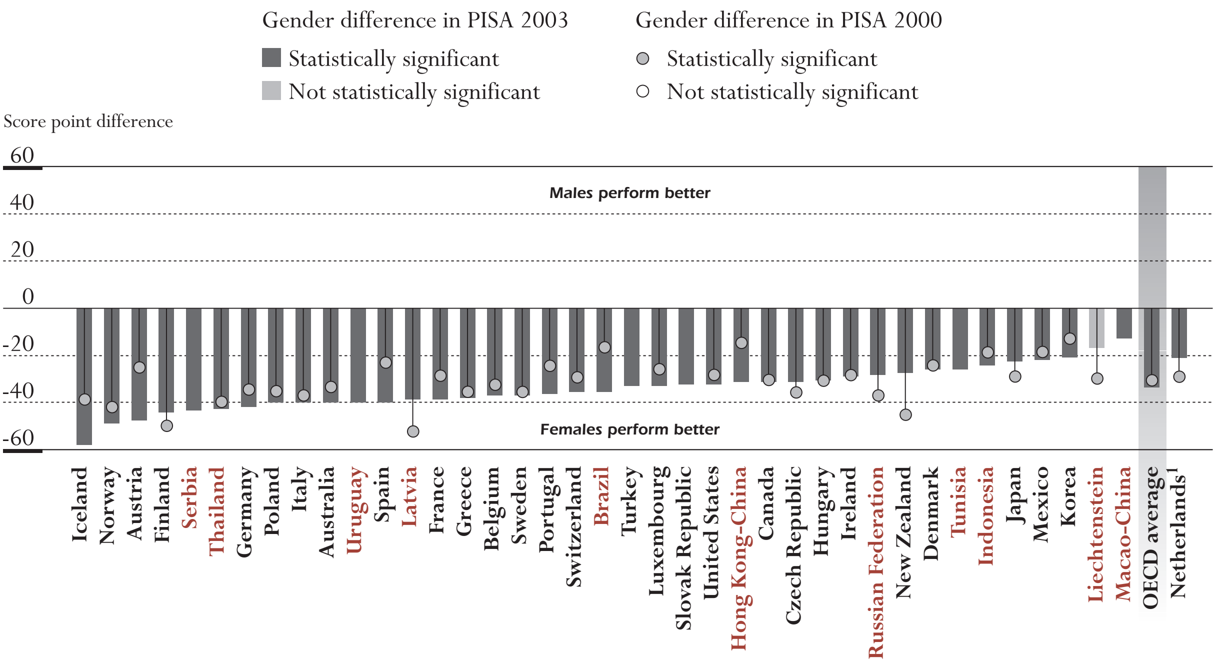
Gender differences in reading literacy

Females perform better at reading than males, but to different degrees across countries.

Figure 6.6 shows differences in performance between males and females for reading in PISA 2000 and PISA 2003 (see also Table 6.3 and Table 5.1 in OECD, 2001a). The panel shows a similar picture to what was found in 2000. Females have significantly higher average performance in reading in all countries with the exception of Liechtenstein, with an average OECD gap in reading of 34 score points, equivalent to half a proficiency level (see Chapter 2 and OECD, 2001a). There is variation across countries in the magnitude of this difference: for example, at least 40 score points separate females from males in reading performance in Austria, Finland, Germany, Iceland, Norway and Poland and in the partner countries Serbia and Thailand. The gender difference is particularly high in Iceland where it reaches 58 score points.

Figure 6.6 ■ **Gender differences in reading performance in PISA 2003 and PISA 2000**

Differences in PISA scale scores



1. The 2000 response rate in the Netherlands was too low to ensure comparability (see Annex A3, OECD 2001a).

Source: OECD PISA 2003 database, Table 6.3; OECD PISA 2000 database, Table 5.1a (OECD 2001a).

For these countries, the average score for females falls within Level 3 while the average score for males falls within Level 2, with the exception of Finland where females score on average within Level 4 while males score on average within Level 3.

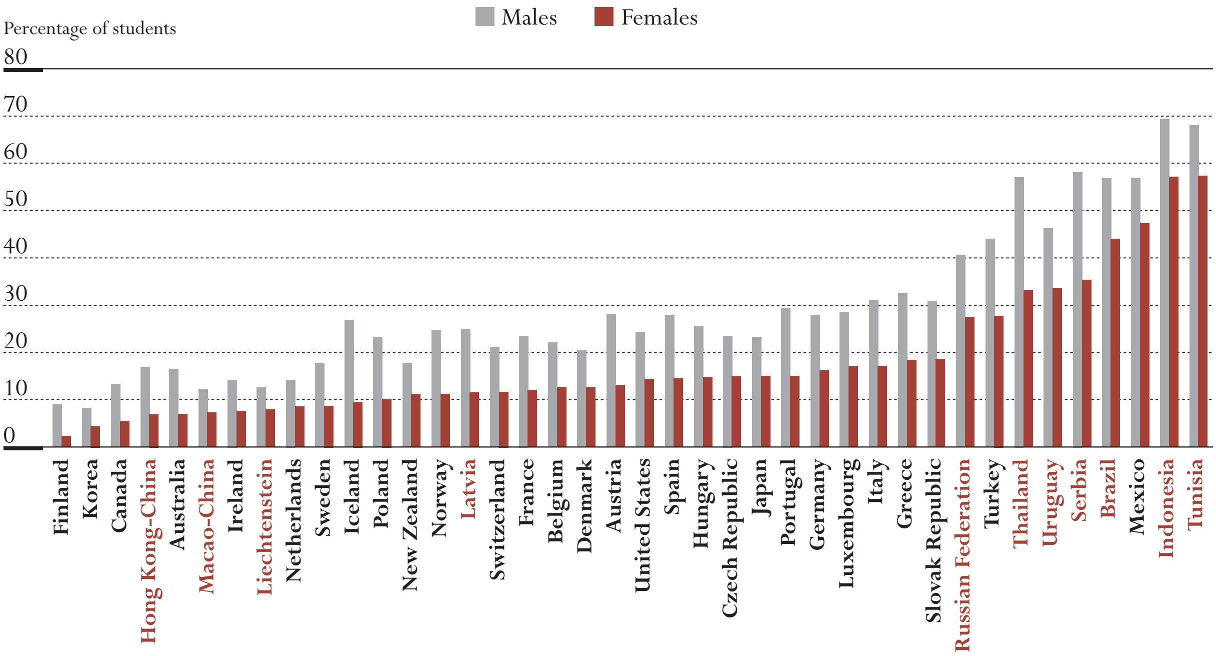
The better performance of females in reading and males in mathematics (see Chapter 2) are consistent with results found in other studies for similar age groups.

When the gender gap found in PISA 2003 is compared with the gap found in PISA 2000, they are in general consistent. However, there are some exceptions.

In many countries males are far more likely than females to be among the lowest performers.

One way to understand the gender differences is to examine the extremes of the distribution. Previous studies have also shown that gender differences in performance increase towards the extremes of the distribution of performance and the large gender differences among students with the lowest levels of performance is of concern to policy makers. In all participating countries, except for the partner countries Liechtenstein and Macao-China, males are significantly more likely than females to be among the lowest-performing students. In 12 OECD countries males are at least twice as likely than females to score below 400 score points (*i.e.,* one standard deviation below the OECD average) and in Finland and Iceland they are three times or more as likely (Table 6.4).

Figure 6.7 ■ **Proportion of males and females among the lowest performers on the reading scale**
Percentage of males and females at or below Level 1



Source: OECD PISA 2003 database, Table 6.5.

Figure 6.7 shows the percentages of males and females scoring at or below Level 1 in reading (Table 6.5). In Iceland, while 10 per cent of females score at or below Level 1, the percentage of males is 27 per cent. In the partner countries Serbia and Thailand, there are at least 20 per cent more males than females at or below Level 1. Among the OECD countries, the smallest differences between the percentages of males and females at lower levels of performance are found in Korea and the Netherlands and in the partner countries, these are found in Liechtenstein and Macao-China.

HOW SCIENCE PERFORMANCE IS MEASURED IN PISA

The science assessment emphasises the application of knowledge…

The emphasis of the PISA 2003 assessment of science is on the application of science knowledge and skills in real-life situations, as opposed to testing particular curricular components. Scientific literacy is defined as the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity.

This definition is based on three dimensions: scientific knowledge or concepts, scientific processes and the situations or context in which the knowledge and processes are assessed.

With the limited assessment time that was available for science in 2003, it was not possible to assess all areas of scientific knowledge, so a sample of concepts

was assessed. The selection of these concepts from the major scientific fields of physics, chemistry, biological science, and earth and space science was guided by a number of principles. First, the knowledge assessed should be relevant to real life situations. Second, the knowledge assessed should have some enduring relevance to life over the next decade at least. Third, the knowledge required to successfully answer a PISA science item should be related to some important scientific process – that is, it should not be an isolated recall of a piece of information.

… focusing on a selection of concepts that are central to science, of enduring relevance and important to real life.

Three main scientific processes are part of the PISA assessment in 2003. The first of these is describing, explaining and predicting scientific phenomena – important facets of the scientific process. Students were given tasks that involved recognising phenomena, giving explanations and making considered judgements as to the impact of these phenomena. The second is understanding scientific investigation, which involves being able to recognise questions and problems that could be solved using scientific methods and what evidence may be needed to achieve this, and may also involve an understanding of the variables that need to be measured and controlled in an experiment. In addition, students were assessed on their ability to communicate these ideas. The third is interpreting scientific evidence and conclusions, which is concerned with the use of scientific findings as evidence for a diverse range of claims and conclusions. Through the media, students are constantly coming into contact with claims made by advertisers, proponents of change and commentators who use scientific evidence as a justification.

Students were required to recognise and explain scientific phenomena, understand scientific investigation and interpret evidence…

The third main aspect of the assessment of science in PISA is a consideration of the areas of application. For PISA 2003 these are science in life and health, science in the earth and environment, and science in technology. The range of assessment tasks includes problems that affect people as individuals (such as food and energy use), as members of a local community (such as the location of a power station) or as world citizens (such as global warming).

….with tasks drawn from a range of scientific situations.

Following PISA 2000, two units, which contained eight items, were released to give an indication of the type of problems that students were encountering (OECD, 2002c). These items were replaced with newly created ones which underwent an extensive field trial process to ensure they had similar levels of difficulty as the released items. A sufficient number of items was retained to allow linking to occur between the assessments carried out different times.

The 2003 science assessment overlapped with that used in 2000…

Like performance in reading literacy, performance in science was marked in PISA 2000 on a single scale with an average score of 500 score points and a standard deviation of 100 score points. Approximately two-thirds of students across OECD countries scored between 400 and 600 score points. The same scale was used for the PISA 2003 science assessment. The scale measures students' ability to use scientific knowledge (understanding of scientific concepts), to recognise scientific questions and to identify what is involved in scientific investigations (understanding of the nature of scientific investigation), to relate scientific data to claims and conclusions (use of scientific evidence) and to communicate these aspects of science.

…and results were reported on the same scale.

Figure 6.8 ■ **A sample of science items used in PISA: Unit DAYLIGHT**

DAYLIGHT

Read the following information and answer the questions that follow.

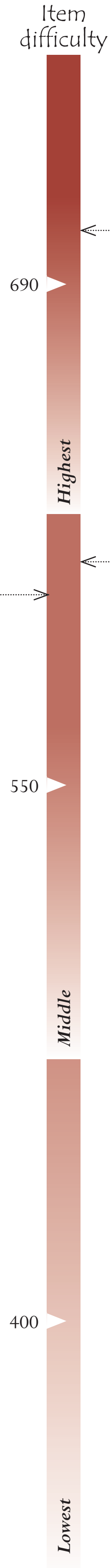
Today, as the Northern Hemisphere celebrates its longest day, Australians will experience their shortest.

In Melbourne,* Australia, the sun will rise at 7:36 am and set at 5:08 pm, giving nine hours and 32 minutes of daylight.

Compare today to the year's longest day in the Southern Hemisphere, expected on 22 December, when the sun will rise at 5:55 am and set at 8:42 pm, giving 14 hours and 47 minutes of daylight.

The President of the Astronomical Society, Mr Perry Vlahos, said the existence of changing seasons in the Northern and Southern Hemispheres was linked to the Earth's 23-degree tilt.

*Melbourne is a city in Australia at a latitude of about 38 degrees south of the equator.



DAYLIGHT

QUESTION 1

Which statement explains why daylight and darkness occur on earth?

A. The earth rotates on its axis.

B. The sun rotates on its axis.

C. The earth's axis is tilted.

D. The earth revolves around the sun.

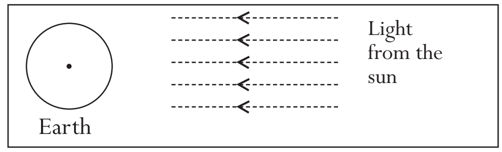
Score 1 (592)
The correct answer is option A.

This is a multiple-choice item that requires students to be able to relate the rotation of the earth on its axis to the phenomenon of day and night and to distinguish this from the phenomenon of the seasons, which arises from the tilt of the axis of the earth as it revolves around the sun. All four alternatives given are scientifically correct.

QUESTION 2

In the Figure light rays from the sun are shown shining on the earth.

Figure: light rays from the sun



Suppose it is the shortest day in Melbourne.

Show the earth's axis, the Northern Hemisphere, the Southern Hemisphere and the Equator on the figure. Label all parts of your answer.

Score 2 (720)
Answers which include a diagram with the Equator tilted towards the sun at an angle between 10° and 45° and the earth's axis tilted towards the sun within the range 10° and 45° from vertical, and the Northern and or Southern Hemispheres correctly labelled (or one only labelled, the other implied).

Score 1 (667)
Answers which include a diagram with:

- the angle of tilt of earth's axis between 10° and 45°, the Northern and/or Southern Hemispheres correctly labelled (or one only labelled, the other implied), but angle of tilt of the Equator not between 10° and 45°; or the Equator missing.
- the angle of tilt of the Equator between 10° and 45°, the Northern and/or Southern Hemispheres correctly labelled (or one only labelled, the other implied), but angle of tilt of axis not between 10° and 45°; or axis missing.
- the angle of tilt of the Equator between 10° and 45°, and angle of tilt of axis between 10° and 45°, but the Northern and Southern Hemispheres not correctly labelled (or one only labelled, the other implied, or both missing).

This is an open-response item that requires students to create a conceptual model in the form of a diagram showing the relationship between the rotation of the earth on its tilted axis and its orientation to the sun on the shortest day for a city in the southern hemisphere. In addition they had to include in this diagram the position of the equator at a 90-degree angle to the tilted axis. Full credit is obtained if the students correctly place and label all three significant elements – the hemispheres, the tilted axis and the equator. Partial credit is given for a diagram with two of the three elements correctly placed and labelled.

Figure 6.9 ■ **A sample of science items used in PISA: Unit CLONING**

CLONING

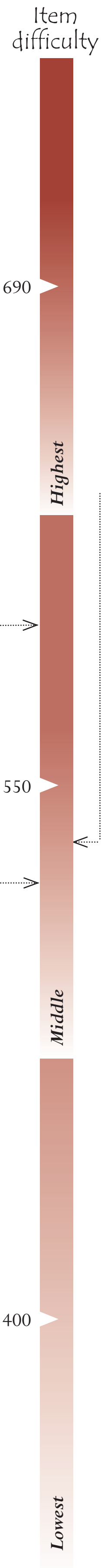
Read the newspaper article and answer the questions that follow.

A copying machine for living beings?

Without any doubt, if there had been elections for the animal of the year 1997, Dolly would have been the winner! Dolly is a Scottish sheep that you see in the photo. But Dolly is not just a simple sheep. She is a clone of another sheep. A clone means: “a copy”. Cloning means: “copying from a single master copy“. Scientists succeeded in creating a sheep (Dolly) that is identical to a sheep that functioned as a master copy.

It was the Scottish scientist Ian Wilmut who designed the “copying machine“ for sheep. He took a very small piece from the udder of an adult sheep (sheep 1). From that small piece he removed the nucleus, then he transferred the nucleus into the egg-cell of another (female) sheep (sheep 2). But first he removed from that egg-cell all the material that would have determined sheep 2 characteristics in a lamb produced from that egg-cell. Ian Wilmut implanted the manipulated egg-cell of sheep 2 into yet another (female) sheep (sheep 3). Sheep 3 became pregnant and had a lamb: Dolly.

Some scientists think that within a few years it will be possible to clone people as well. But many governments have already decided to forbid the cloning of people by law.



CLONING

Question 1

Which sheep is Dolly identical to?

A. Sheep 1

B. Sheep 2

C. Sheep 3

D. Dolly's father

Score 1 (494)

The correct answer is option A.

This is a multiple-choice question item that assesses the students' understanding of the process by which the cloning takes place. This is described in detail in the text, and the students are required to carefully read this text to extract the information required. They need to know that the nucleus of the cell contains the material that will determine the characteristics of the off-spring.

QUESTION 2

In line 14 the part of the udder that was used is described as "a very small piece". From the article text you can work out what is meant by "a very small piece".

That "very small piece" is

A. a cell.

B. a gene.

C. a cell nucleus.

D. a chromosome.

Score 1 (572)

The correct answer is option A.

This is a multiple-choice item that requires the students to demonstrate an understanding of the structure of cells.

QUESTION 3

In the last sentence of the article it is stated that many governments have already decided to forbid the cloning of people by law.

Two possible reasons for this decision are mentioned below.

Are these reasons scientific reasons?

Circle either "Yes" or "No" for each.

Reason:	Scientific?
Cloned people could be more sensitive to certain diseases than normal people.	Yes/No
People should not take over the role of a Creator.	Yes/No

Score 1 (507)

Answers which indicate Yes, No, in that order.

This is a complex multiple-choice item that requires students to show that they can distinguish between statements that are scientifically based and those that are not. One of the aspects of the PISA scientific literacy framework is the notion that students understand scientific investigation and reasoning. The question poses two reasons why governments might forbid human cloning. One of the reasons is concerned with the fact that cloned people might be more susceptible to disease (a reason that could be said to be "scientific"), while the other is statement that people should not take on the role of a Creator (a valid reason for many people, but one which cannot be said to be "scientific"). Full credit is obtained for correctly labelling both statements.

More difficult tasks involve more complex concepts and greater skill requirements, and demand more sophisticated scientific knowledge.

The increasing difficulty of tasks along the scale involves the complexity of the concepts used, the amount of data given, the chain of reasoning required and the precision required in communication. In addition, the level of difficulty is influenced by the context of the information, the format and the presentation of the question. The tasks in PISA require scientific knowledge involving (in ascending order of difficulty): recall of simple scientific knowledge or common scientific knowledge or data; the application of scientific concepts or questions and a basic knowledge of investigation; the use of more highly developed scientific concepts or a chain of reasoning; and knowledge of simple conceptual models or analysis of evidence in order to try out alternative approaches.

Science is not rated at proficiency levels, but it is possible to define characteristics of difficult, medium and easy scientific tasks.

Unlike for reading and mathematics (see Chapter 2), the science scale cannot yet be defined in terms of proficiency levels. This will only be possible from 2006 onwards, when science becomes the main focus of the PISA assessment for the first time and when a full instrument for measuring and reporting science will be developed. However, the criteria for harder and easier tasks can still be described in relation to items associated with different points on the science scale.

- Towards the top end of the science scale (around 690 score points) students are generally able to create or use conceptual models to make predictions or give explanations; to analyse scientific investigations in order to grasp, for example, the design of an experiment or to identify an idea being tested; to compare data in order to evaluate alternative viewpoints or differing perspectives; and to communicate scientific arguments and/or descriptions in detail and with precision.
- At around 550 score points, students are typically able to use scientific concepts to make predictions or provide explanations; to recognise questions that can be answered by scientific investigation and/or identify details of what is involved in a scientific investigation; and to select relevant information from competing data or chains of reasoning in drawing or evaluating conclusions.
- Towards the lower end of the scale (around 400 score points), students are able to recall simple factual scientific knowledge (*e.g.,* names, facts, terminology, simple rules); and to use common scientific knowledge in drawing or evaluating conclusions.

A full description of the conceptual framework underlying the PISA assessment of science is provided in *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Kowledge and Skills* (OECD, 2003e).

The tasks used for the assessment of science in PISA are quite diverse. Figure 6.8 and Figure 6.9 show examples of the science tasks used in PISA 2003, along with a description of the criteria used to mark students' answers. A more complete set of sample tasks can be found at *www.pisa.oecd.org*. The science assessment was comprised of 35 items divided into 13 units from which 25 items from 10 units were the same as the ones used in 2000 (see Annex A6, Table A6.3 for the breakdown of the items by the various aspects of the framework).

The sample unit *Daylight* provides verbal information on the variation in the length of daylight between the Northern and Southern hemispheres (Figure 6.8). The change of seasons in these hemispheres is also related to the tilt of the earth's axis.

The stimulus for the sample unit, *Cloning*, features an extract from a newspaper article and a photograph of Dolly, the first sheep to be cloned (Figure 6.9). The questions that follow are probing the students' knowledge of the structure of animal cells and scientific methods of investigation.

When taken together, these science units help to illustrate the underlying understanding of science that PISA has adopted in its framework as scientific literacy, in particular the ability to use science knowledge to give explanations.

STUDENT PERFORMANCE IN SCIENCE

The mean performances of countries in science

On average, students did as well in science in 2003 as in 2000, but their results were slightly more spread out.

As previously described in Chapter 2 for the case of mathematics and earlier in this Chapter for reading, the average scores of countries provide an indication of the overall level of performance, keeping in mind that mean scores provide an incomplete picture of performance. As with reading, the outcomes for science are based on the science scale that was developed for PISA 2000 and which had a mean of 500 and a standard deviation of 100. Figure 6.10 shows average performance on the science scale (Table 6.6). The PISA 2003 results include 29 OECD countries – the Slovak Republic and Turkey joined PISA in 2003 and the Netherlands met all technical standards in 2003, while the United Kingdom has been excluded from the results as it failed to reach the technical standards required by PISA 2003.

Four countries had the highest performance and their averages are indistinguishable.

When the 25 OECD countries for which comparable data are available for both the PISA 2000 and 2003 assessments are compared jointly, it is clear that the average performance has remained unchanged (Figure 6.10).[6] However, mainly because of the inclusion of new countries in 2003, the overall OECD mean for science is now 496 score points and the standard deviation is 105 score points.

The gap in performance between the highest and the lowest performing OECD countries is 143 points. That is, while the average scores of the highest performing countries of Finland and Japan is 548 or about half a standard deviation above the OECD average, Mexico's average score of 405 score points is almost one standard deviation below the OECD average.

Finland and Japan have the highest mean scores and rank between first and third on the science scale, but their performance is not statistically significantly different from that in Korea and the partner country Hong Kong-China, who both rank between second and fourth. Other OECD countries that show mean performance in science higher than the OECD average are Australia, Belgium, Canada, the Czech Republic, France, Ireland, the Netherlands, New Zealand, Sweden, Switzerland and among the partner countries Liechtenstein and Macao-China. Countries with performance not statistically different from the OECD average are Germany, Hungary, Poland and the Slovak Republic.[7]

Figure 6.10 ■ Multiple comparisons of mean performance on the science scale

Science scale	Mean	S.E.	Finland	Japan	Hong Kong-China	Korea	Liechtenstein	Australia	Macao-China	Netherlands	Czech Republic	New Zealand	Canada	Switzerland	France	Belgium	Sweden	Ireland	Hungary	Germany	Poland	Slovak Republic	Iceland	United States	Austria	Russian Fed.	Latvia	Spain	Italy	Norway	Luxembourg	Greece	Denmark	Portugal	Uruguay	Serbia	Turkey	Thailand	Mexico	Indonesia	Brazil	Tunisia
	Mean		548	548	539	538	525	525	525	524	523	521	519	513	511	509	506	505	503	502	498	495	495	491	491	489	489	487	486	484	483	481	475	468	438	436	434	429	405	395	390	385
		S.E.	(1.9)	(4.1)	(4.3)	(3.5)	(4.3)	(2.1)	(3.0)	(3.1)	(3.4)	(2.4)	(2.0)	(3.7)	(3.0)	(2.5)	(2.7)	(2.7)	(2.8)	(3.6)	(2.9)	(3.7)	(1.5)	(3.1)	(3.4)	(4.1)	(3.9)	(2.6)	(3.1)	(2.9)	(1.5)	(3.8)	(3.0)	(3.5)	(2.9)	(3.5)	(5.9)	(2.7)	(3.5)	(3.2)	(4.3)	(2.6)
Finland	548	(1.9)		●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Japan	548	(4.1)	●		●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hong Kong-China	539	(4.3)	●	●		●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Korea	538	(3.5)	●	●	●		●	▲	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Liechtenstein	525	(4.3)	▼	▼	●	●		●	●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Australia	525	(2.1)	▼	▼	●	▼	●		●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Macao-China	525	(3.0)	▼	▼	●	●	●	●		●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Netherlands	524	(3.1)	▼	▼	●	●	●	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Czech Republic	523	(3.4)	▼	▼	●	●	●	●	●	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
New Zealand	521	(2.4)	▼	▼	▼	▼	●	●	●	●	●		●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Canada	519	(2.0)	▼	▼	▼	▼	●	●	●	●	●	●		●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Switzerland	513	(3.7)	▼	▼	▼	▼	●	●	●	●	●	●	●		●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
France	511	(3.0)	▼	▼	▼	▼	●	▼	●	●	●	●	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Belgium	509	(2.5)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●		●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Sweden	506	(2.7)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Ireland	505	(2.7)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hungary	503	(2.8)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●		●	●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Germany	502	(3.6)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●		●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Poland	498	(2.9)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●		●	●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Slovak Republic	495	(3.7)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Iceland	495	(1.5)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●		●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
United States	491	(3.1)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●		●	●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Austria	491	(3.4)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●		●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Russian Fed.	489	(4.1)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●		●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Latvia	489	(3.9)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●		●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Spain	487	(2.6)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Italy	486	(3.1)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Norway	484	(2.9)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	●	●	●	●	●	●		●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Luxembourg	483	(1.5)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	●	●	●	●	●	●	●		●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Greece	481	(3.8)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	●	●	●	●	●	●	●	●		●	●	▲	▲	▲	▲	▲	▲	▲	▲
Denmark	475	(3.0)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●		●	▲	▲	▲	▲	▲	▲	▲	▲
Portugal	468	(3.5)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		▲	▲	▲	▲	▲	▲	▲	▲
Uruguay	438	(2.9)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼		●	●	●	▲	▲	▲	▲
Serbia	436	(3.5)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●		●	●	▲	▲	▲	▲
Turkey	434	(5.9)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	▲	▲	▲	▲
Thailand	429	(2.7)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●		▲	▲	▲	▲
Mexico	405	(3.5)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼		●	●	▲
Indonesia	395	(3.2)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●		●	●
Brazil	390	(4.3)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●
Tunisia	385	(2.6)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	

Range of rank*

		Finland	Japan	Hong Kong-China	Korea	Liechtenstein	Australia	Macao-China	Netherlands	Czech Republic	New Zealand	Canada	Switzerland	France	Belgium	Sweden	Ireland	Hungary	Germany	Poland	Slovak Republic	Iceland	United States	Austria	Russian Fed.	Latvia	Spain	Italy	Norway	Luxembourg	Greece	Denmark	Portugal	Uruguay	Serbia	Turkey	Thailand	Mexico	Indonesia	Brazil	Tunisia
OECD countries	Upper rank	1	1		2		4		4	4	4	6	7	9	9	10	10	11	11	14	15	16	17	16			19	19	20	22	21	25	26			28		29			
	Lower rank	2	3		3		7		8	8	8	9	13	13	13	15	15	16	17	19	21	19	23	23			24	25	25	25	26	27	27			28		29			
All countries	Upper rank	1	1	2	2	5	5	5	5	5	6	8	10	12	12	13	13	14	14	17	18	19	20	19	20	20	22	22	24	26	25	30	31	33	33	33	34	37	38	38	39
	Lower rank	3	3	4	4	11	10	10	11	11	11	12	15	16	16	18	18	19	21	22	25	23	27	28	30	29	29	30	30	30	31	32	32	35	36	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.

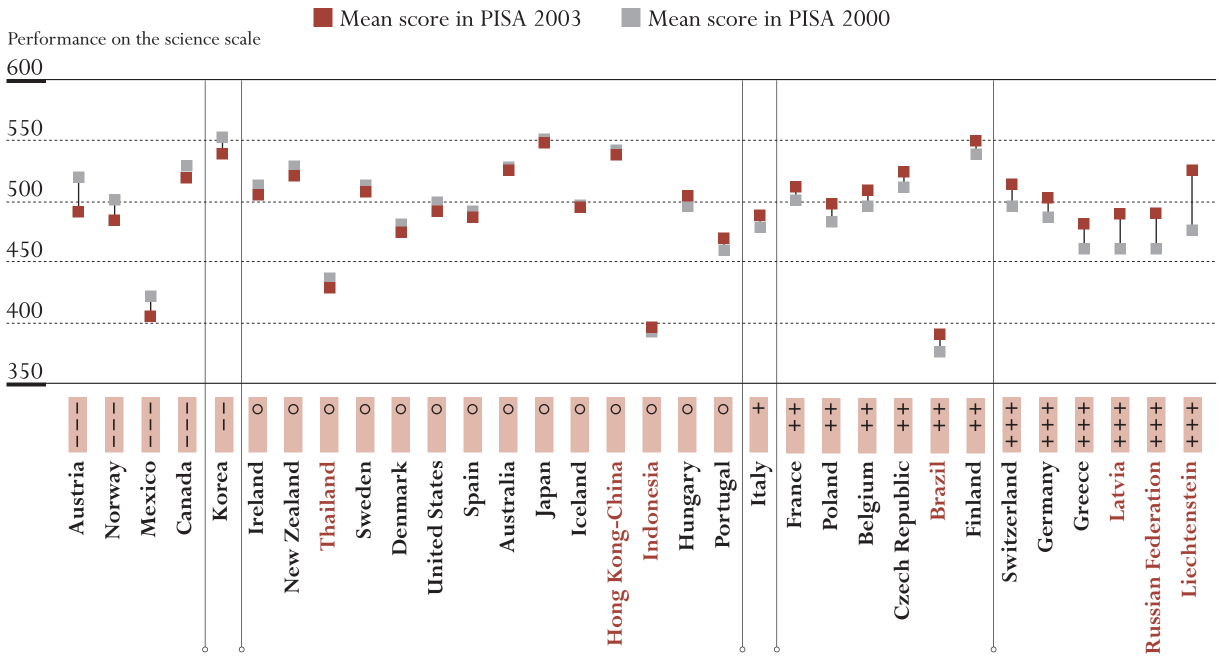
Without the Bonferroni adjustment:
- Mean performance statistically significantly higher than in comparison country
- No statistically significant difference from comparison country
- Mean performance statistically significantly lower than in comparison country

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.

Figure 6.11 ■ **Differences in mean scores between PISA 2003 and PISA 2000 on the science 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 database, Table 6.6; OECD PISA 2000 database, Table 3.3 (OECD 2001a).

Differences in science performance between PISA 2000 and PISA 2003

Results for the two science surveys should be compared cautiously.

Most of the science items that were used for assessment in 2000 were also used in 2003. This meant that links could be made with any of the new items that were used and, consequently, changes from 2000 to 2003 could be considered. Figure 6.11 shows science scores for PISA 2000 and 2003 for the countries and indicates differences in performance between the two assessments. However, as explained before, such differences need to be interpreted with caution.

Some countries showed improvement, most often driven by higher-ability students…

Thirteen countries, among them nine OECD countries, showed statistically significant increases in overall performance from PISA 2000 to PISA 2003 as indicated by the mean score. These include Belgium, the Czech Republic, Finland, France, Germany, Greece, Italy, Poland and Switzerland as well as the partner countries Brazil, Latvia, Liechtenstein and the Russian Federation. Figure 6.12 shows the differences within each country at the various percentile levels. In Belgium, the Czech Republic, Finland, France, Germany, Italy, Poland and the partner country Brazil the increases tended to be driven by improvements in the upper half of the performance distribution (the 75th, 90th and 95th percentiles), *i.e.,* the better performing students became better.

Figure 6.12 ■ **Comparisons between PISA 2003 and PISA 2000 in science**

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

Source: OECD PISA 2003 database, Table 6.6 and OECD PISA 2000 database, Table 3.3 (OECD 2001a).

. . . while science performance fell in a smaller number of countries, most often pulled down by lower-ability students.

Five countries showed a significant decline in performance, namely Austria, Canada, Korea, Mexico and Norway. For Korea, while the top performing 5 per cent of students showed higher performance in 2003, the 25 per cent lowest-performing students performed markedly lower, dragging overall performance down. The picture is similar for Japan and Sweden, but with no difference in average performance.

Gender differences in science

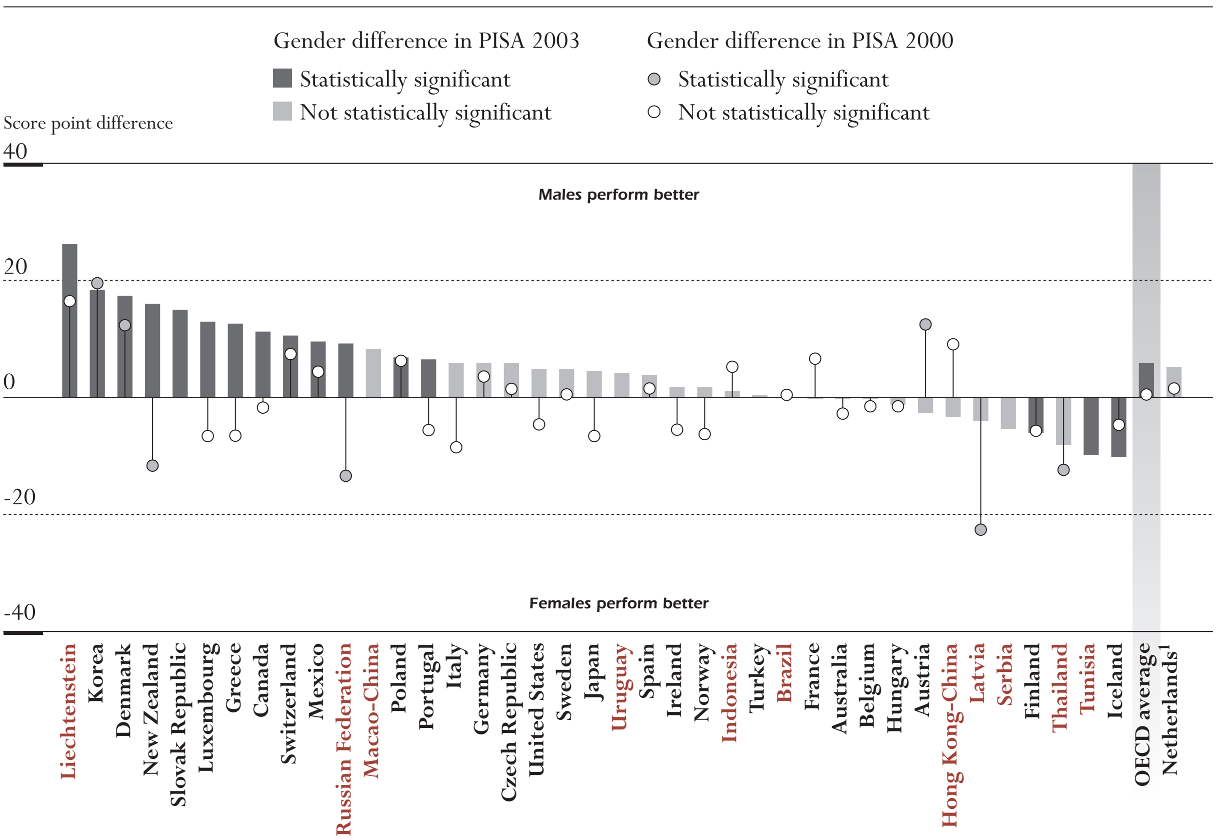
Science showed the smallest average gender differences among all content areas assessed.

As in PISA 2000, science showed the smallest average gender differences among all content areas assessed (Table 6.7 and Figure 6.13), with an OECD average difference between males and females of six score points in favour of males. Statistically significant differences in favour of males are found in Canada, Denmark, Greece, Korea, Luxembourg, Mexico, New Zealand, Poland, Portugal, Slovak Republic and Switzerland and in the partner countries Liechtenstein and the Russian Federation. On the other hand, females in Finland, Iceland and the partner country Tunisia outperform males.

Another way of looking at the distribution of scores is to examine the percentage of students scoring below 400 score points, *i.e.,* one standard deviation below the OECD mean and the percentage of students scoring above 600 score points – that is, one standard deviation above the OECD mean. This is particularly useful in science given that performance has not been classified by proficiency levels. With around two-thirds of the students scoring between 400 and 600 score points, around one-sixth of students perform at each of these extremes.

As expected from the previous analysis in PISA 2000 which showed a minimal level of gender differences in the results of science, there are also very small differences between the percentage of males and females scoring below 400 score points (less than 5 per cent in either direction for the OECD countries). The same is true for students scoring above 600 score points (Table 6.8).

Figure 6.13 ■ **Gender differences in science performance in PISA 2003 and PISA 2000**
Differences in PISA scale scores



1. The response rate in the Netherlands in 2000 is to low to ensure comparability (see Annex A3, OECD, 2001a).
Source: OECD PISA 2003 database, Table 6.7; OECD (2001a), Table 5.1a.

IMPLICATIONS FOR POLICY

Reading

The results for PISA 2000 show wide differences between countries in the knowledge and skills of 15-year-olds in reading literacy. Differences between countries represent, however, only a fraction of overall variation in student performance, with differences within countries being on average about ten times as great as the variation between country means.

The persistence of a small but significant minority of students unable to perform even simple reading tasks remains of concern…

Catering for such a diverse client base and narrowing the gaps in student performance represents formidable challenges for all countries: An average of 8 per cent of 15-year-olds reach the highest reading level in PISA, demonstrating the ability to complete sophisticated reading tasks, to show detailed understanding of texts and the relevance of their components, and to evaluate information critically and build hypotheses drawing on specialised knowledge. At the other end of the scale, an average of 8 per cent of students do not reach proficiency Level 1. They fail to demonstrate routinely the most basic knowledge and skills that PISA seeks to measure. These students may still be able to read in a technical sense, but they show serious difficulties in applying reading literacy as a tool to advance and extend their knowledge and skills in other areas. Although the proportion of these students is below 2 per cent in three countries, including two OECD countries, and exceeds 10 per cent in only three OECD and seven partner countries, the existence of a small but significant minority of students who, near the end of compulsory schooling, lack the foundation of literacy skills needed for further learning, must be of concern to policy makers seeking to make lifelong learning a reality for all. This is so, in particular, in the face of mounting evidence that continuing education and training beyond school tend to reinforce rather than to mitigate skill differences resulting from unequal success in initial education.

…as does the nearly one in five who can only perform the simplest tasks.

Adding to this proportion of students not reaching Level 1 those who perform only at Level 1, namely those who are capable only of completing the most basic of reading tasks, such as locating a simple piece of information, identifying the main theme of a text or making a simple connection with everyday knowledge, brings the proportion of low performers at or below Level 1 to an average of 19 per cent across OECD countries. Parents, educators, and policy makers in systems with large proportions of students performing at or below Level 1 need to recognise that significant numbers of students are not benefiting sufficiently from available educational opportunities and are not acquiring the necessary knowledge and skills to do so effectively in their further school careers and beyond.

In some higher-performing countries, a wide distribution can be of concern even if most students do comparatively well.

Wide variation in student performance does not, however, always mean that a large part of the student population will have a low level of reading literacy. In fact, in some countries with high average performance, the 25th percentile on the combined reading literacy scale lies well within proficiency Level 2, indicating that students at the 25th percentile are doing reasonably well by international comparative standards. Nevertheless, the variation in the distribution of

student performance in these countries suggests that the students at the 25th percentile may be performing substantially below expected benchmarks of good performance in the countries in question.

To what extent is the observed variation in student performance on the PISA 2003 assessments a reflection of the innate distribution of students' abilities and thus a challenge for education systems that cannot be influenced directly by education policy? The analysis in this chapter has shown not only that the magnitude of within-country disparities in reading literacy varies widely between countries but also that wide disparities in performance are not a necessary condition for a country to attain a high level of overall performance. Although more general contextual factors need to be considered when such disparities are compared between countries, public policy may therefore have the potential to make an important contribution to providing equal opportunities and equitable learning outcomes for all students. Showing that countries differ not just in their mean performance, but also in the extent to which they are able to close the gap between the students with the lowest and the highest levels of performance and to reduce some of the barriers to equitable distribution of learning outcomes is an important finding which has direct relevance for policy makers.

The success of some countries in containing student disparities while achieving high overall performance suggests that education policy can make a difference.

Science

In an increasingly technological world, literacy is not just about reading, but citizens also need to be scientifically literate. Scientific literacy is important for understanding environmental, medical, economic and other issues that confront modern societies, which rely heavily on technological and scientific advances. Further, the performance of a country's best students in scientific subjects may have implications for the part which that country will play in tomorrow's advanced technology sector, and for its general international competitiveness. Conversely, deficiencies in mathematical and scientific literacy can have negative consequences for individuals' labour-market and earnings prospects and for their capacity to participate fully in society.

Scientific literacy is today important for individuals and for society…

Addressing the increasing demand for scientific skills requires excellence throughout education systems, and it is important to monitor how well countries provide young adults with fundamental skills in this area. However, the wide disparities in student performance on the scientific literacy scale that emerge from the analysis in this chapter suggest that this remains still a remote goal and that countries need to serve a wide range of student abilities, including those who perform exceptionally well but also those most in need.

…and countries need to be better at spreading scientific skills to more students.

Gender difference in science, in which males have often been more proficient in past assessments, tend to be much smaller than the difference in favour of females in reading. In fact, in science there is no clear pattern of gender differences, and in most countries gender differences are small. Although it will take time for these results to translate into corresponding participation patterns in higher education as well as occupational structures, this is an encouraging signal.

It is encouraging that gender differences in science are now small.

PISA will assess science performance more thoroughly in 2006.

The 2006 PISA assessment, which will put the main focus on the knowledge, skills and attitudes of 15-year-olds towards science, will reveal to what extent countries are further progressing towards raising science performance, fostering equity in learning opportunities and, perhaps most important of all, developing positive attitudes and dispositions among young adults towards scientific subjects and careers.

Notes

1. For the 25 countries with comparable data in 2000 and 2003, the average performance in 2000 was 501 score points, while the average performance in 2003 was 497 score points. Because of sampling errors and errors associated with the link between the two assessments, the difference is not statistically significant.
2. 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.
3. Comparisons of a particular country average score with the OECD average are based on a recomputed OECD average that excludes the data from the country in question. This is done to avoid dependency between the two averages.
4. See Annex A8 for an explanation of the methodology underlying the link between the PISA 2000 and PISA 2003 assessments.
5. In Luxembourg, the assessment conditions were changed in substantial ways between the 2000 and 2003 assessments in order to reduce linguistic barriers for students. For this reasons, results cannot be compared between 2000 and 2003.
6. For the 25 countries with comparable data in 2000 and 2003, average performance was 501 score points in both the 2000 and the 2003 assessments.
7. Comparisons of a particular country's average score with the OECD average are based on a recomputed OECD average that excludes the data from the country in question. This is done to avoid dependency between the two averages.

REFERENCES

Artelt, C. (2000), *Strategisches Lernen,* Waxmann, Münster.

Bandura, A. (1994), *Self-Efficacy: The Exercise of Control,* Freeman, New York.

Beaton, A.E., **M.O. Martin**, **I.V.S. Mullis**, **E.J. Gonzalez**, **T.A. Smith** and **D.L. Kelly** (1996), *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS),* Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, Chestnut Hill, M.A.

Bempechat, J., **N.V. Jimenez** and **B.A. Boulay** (2002), "Cultural-Cognitive Issues in Academic Achievement: New Directions for Cross-National Research", in A.C. Porter and A. Gamoran (eds.), *Methodological Advances in Cross-National Surveys of Educational Achievement,* National Academic Press, Washington, D.C.

Boekaerts, M. (1999), "Self-regulated learning: Where we are today", *International Journal of Educational Research*, Vol. 31, Issue 6, Elsevier Ltd., pp.445-475.

Brown, A.L., **J.D. Bransford**, **R.A. Ferrara** and **J.C. Campione** (1983), "Learning, remembering and understanding", in J. H. Flavell and E. M. Markman (eds.), *Handbook of Child Psychology, Cognitive Development,* New York, Wiley, pp. 77-166.

Datcher, L. (1982), "Effects of Community and Family Background on Achievement", *Review of Economics and Statistics,* Vol. 64, No. 1, The MIT Press, Cambridge, M.A., pp. 32-41.

Deci, E.L. and **R.M. Ryan** (1985), *Intrinsic Motivation and Self-Determination in Human Behavior,* Plenum Press, New York.

Eccles, J. S. (1994), "Understanding women's educational and occupational choice: Applying the Eccles et al. model of achievement-related choices", *Psychology of Women Quarterly,* Vol. 18, Blackwell Publishing, Oxford, pp. 585-609.

Finn, J. (1989), "Withdrawing from school", *Review of Educational Research*, Vol. 59, No. 2, American Educational Research Association , Washington, D.C., pp. 117-142.

Finn, J.D. (1993), *School Engagement & Students At Risk,* National Center for Educational Statistics, Washington, D.C.

Finn, J. and **D.A. Rock** (1997), "Academic success among students at risk for school failure", *Journal of Applied Psychology,* Vol. 82, No. 2, American Psychological Association, Washington, D.C., pp. 221-234.

Flavell, J.H. and **H.M. Wellman** (1977), "Metamemory", in R.V. Kail, Jr. and W. Hagen (eds.), *Perspectives on the Development of Memory and Cognition*, Erlbaum, Hillsdale, N.J., pp. 3-31.

Ganzeboom, H.B.G., **P.M. De Graaf** and **D.J. Treiman** (1992), "A standard international socio-economic index of occupational status", *Social Science Research*, Vol. 21, Issue 1, Elsevier Ltd., pp. 1-56.

Hart, B. and **T.R. Risely** (1995), *Meaningful Differences in the Everyday Experience of Young American Children,* Brookes, Baltimore, M.D.

Hatano, G. (1998), "Comprehension activity in individuals and groups", in M. Sabourin, F. Craik and M. Robert (eds.), *Advances in Psychological Science, Volume 2: Biological and Cognitive Aspects,* Psychology Press/Erlbaum, Hove, pp. 399-417.

Heine, S.J., **Lehman, D.R.**, **Markus, H.R.** and **Kitayama, S.** (1999), "Is there a universal need for positive self-regard?", *Psychological Review,* Vol. 106, No. 4, American Psychological Association, Washington, D.C., pp. 766-794.

Jenkins, P. H. (1995), "School delinquency and school commitment", *Sociology of Education*, Vol. 68, American Sociological Association, Washington, D.C., pp. 221-239.

Johnson, M. K., **R. Crosnoe** and **G.H. Elder** (2001), "Students' attachment and academic engagement: The role of race and ethnicity", *Sociology of Education,* Vol. 74, American Sociological Association, Washington, D.C., pp.318-340.

Lehtinen, E. (1992), "Lern- und Bewältigungsstrategien im Unterricht", in H. Mandl and F.H. Friedrich (eds.), *Lern- und Denkstrategien: Analyse und Intervention,* Hogrefe, Göttingen, pp. 125-149.

Rosenshine, B. and **C. Meister** (1994), "Reciprocal teaching: A review of the research", *Review of Educational Research,* Vol. 64, No. 4, American Educational Research Association, Washington, D.C., pp. 479-531.

Marsh, H.W. (1986), "Verbal and math self-concepts: An internal/external frame of reference model", *American Educational Research Journal,* Vol. 23, No. 1, American Educational Research Association , Washington, D.C., pp. 129-149.

Marsh, H.W. (1993), "The multidimensional structure of academic self-concept: Invariance over gender and age", *American Educational Research Journal,* Vol. 30, No. 4, American Educational Research Association , Washington, D.C., pp. 841-860.

Meece, J.L., **A. Wigfield** and **J.S. Eccles** (1990), "Predictors of math anxiety and its influence on young adolescents' course enrolment intentions and performance in mathematics", *Journal of Educational Psychology,* Vol. 82, No. 1, American Psychological Association, Washington, D.C., pp. 60-70.

OECD (**Organisation for Economic Co-operation and Development**) (1996), *Education at a Glance,* OECD, Paris.

OECD (1997), *Education at a Glance,* OECD, Paris.

OECD (1999a), *Measuring Student Knowledge and Skills – A New Framework for Assessment,* OECD, Paris.

OECD (1999b), *Classifying Educational Programmes: Manual for ISCED-97 Implementation in OECD Countries*, OECD, Paris.

OECD (2000a), *Education at a Glance,* OECD, Paris.

OECD and **Statistics Canada** (2000b), *Literacy in the Information Age,* OECD, Paris and Ottawa.

OECD (2001a), *Knowledge and Skills for Life – First Results from PISA 2000,* OECD, Paris.

OECD (2001b), *Starting Strong - Early Childhood Education and Care,* OECD, Paris.

OECD (2002a*), Manual for the PISA 2000 Database,* OECD, Paris.

OECD (2002b), *Reading for Change – Performance and Engagement across Countries*, OECD, Paris.

OECD (2002c), *Sample Tasks from the PISA 2000 Assessment – Reading, Mathematical and Scientific Literacy,* OECD, Paris.

OECD (2002d), *PISA 2000 Technical Report,* OECD, Paris.

OECD (2002e), *Education Policy Analysis,* OECD, Paris.

OECD (2003a), *Education at a Glance,* OECD, Paris.

OECD (2003b), *Learners for Life – Student Approaches to Learning,* OECD, Paris.

OECD (2003c), *Literacy Skills for the World of Tomorrow – Further Results from PISA 2003,* OECD, Paris.

OECD (2003d), *Student Engagement at School – A Sense of Belonging and Participation,* OECD, Paris.

OECD (2003e), *The PISA 2003 Assessment Framework – Mathematics, Reading, Science and Problem Solving Knowledge and Skills,* OECD, Paris.

OECD (2003f), *Trends in International Migration,* OECD, Paris.

OECD (2004a), *Education at a Glance,* OECD, Paris.

OECD (2004b), *Attracting, Developing and Retaining Effective Teachers,* OECD, Paris.

OECD (2004c), *What Makes School Systems Perform,* OECD, Paris.

OECD (2004d), *Problem Solving for Tomorrow's World – First Measures of Cross-Curricular Competencies,* OECD, Paris.

OECD (forthcoming), *PISA 2003 Technical Report,* OECD, Paris.

OECD and **Statistics Canada** (1995), *Literacy, Economy and Society: Results of the First International Adult Literacy Survey,* OECD, Paris and Ottawa.

Offord, D.R. and **B.G. Waters** (1983), "Socialization and its failure", in M.D. Levine, W.B. Carey, A.C. Crocker and R.T. Gross (eds.), *Developmental-Behavioral Pediatrics,* John Wiley and Sons Inc., New York, pp. 650-682.

Offord, D.R. and **K. Bennett** (1994), "Conduct disorder: Long-term outcomes and intervention effectiveness", *Journal of the American Academy of Child & Adolescent Psychiatry,* Vol. 33, Issue 8, Lippincott Williams & Wilkins, Baltimore, M.D., pp.1069-1078.

Owens, L. and **J. Barnes** (1992), *Learning Preferences Scales,* ACER, Victoria.

Rychen, D.S. and **L.H. Salganik** (eds.) (2002), *Defining and Selecting Key Competencies,* Hogrefe and Huber Publishers, Seattle, W.A.

Schiefele, U., **A. Krapp** and **A. Winteler** (1992), "Interest as a predictor of academic achievement: A meta-analysis of research", in K. A. Renninger, S. Hidi and A. Krapp (eds.), *The Role of Interest in Learning and Development,* Erlbaum, Hillsdale, N.J., pp. 183-212.

Schneider, W. (1996), "Zum Zusammenhang zwischen Metakognition und Motivation bei Lern- und Gedächtnisvorgängen", in C. Spiel, U. Kastner-Koller and P. Deimann (eds.), *Motivation und Lernen aus der Perspektive lebenslanger Entwicklung,* Waxmann, Münster, pp. 121-133.

Schunk, D.H. (1991), *Learning Theories: An Educational Perspective,* Macmillan Publishing Company, New York.

Steen (1990), *On the Shoulders of Giants: New Approaches to Numeracy,* National Academy Press, Washington, DC.

Stanat, P. (2004), "The role of migration background for student performance: an international comparison", paper presented at the 2004 Annual Meeting of the American Educational Research Association (AERA) San Diego, C.A., 12-16 April.

Veenman, M.V. J. and **B.H.A.M. van Hout-Wolters** (2002), "Het meten van metacognitieve vaardigheden", in F. Daems, R. Rymenans and G. Rogiest (eds.), *Onderwijsonderzoek in Nederland en Vlaanderen. Prooceedings van de 29e Onderwijs Research Dagen 2002 te Antwerpen,* Universiteit Antwerpen, Antwerpen, pp. 102-103.

van de Vijver, F. and **K. Leung** (1997), "Methods and data analysis of comparative research", in J. W. Berry, Y., H. Poortinga and J. Pandey (eds.), *Handbook of Cross-Cultural Psychology, Vol. 1 Theory and Method,* Allyn and Bacon, Needham Heights, M.A., pp. 257-300.

Voelkl, K.E. (1995), "School warmth, student participation, and achievement", *Journal of Experimental Education,* Vol. 63, No. 2, HELDREF Publications, Washington, D.C., pp.127-138.

Wang, M., **G. Haertel** and **H. Walberg** (1993), "Toward a knowledge base for school learning", *Review of Educational Research,* Vol. 63, pp. 249-294.

Warm, T.A. (1985), "Weighted maximum likelihood estimation of ability in Item Response Theory with tests of finite length", *Technical Report CGI-TR-85-08,* U.S. Coast Guard Institute, Oklahoma City.

Weinert, F. E. (1994), "Lernen lernen und das eigene lernen verstehen", in K. Reusser and M. Reusser-Weyeneth (eds.), *Verstehen. Psychologischer Prozeß und didaktische Aufgabe,* Huber, Bern, pp. 183-05.

Wigfield, A., **J.S. Eccles** and **D. Rodriguez** (1998), "The development of children's motivation in school context", *Review of Research in Education,* Vol. 23, American Educational Research Association, Washington, D.C. pp. 73-118.

Willms, J. D. (2002), *Vulnerable Children: Findings from Canada's National Longitudinal Survey of Children and Youth,* University of Alberta Press, Edmonton.

Willms, J.D. (2004), "Student Performance and Socio-economic Background", unpublished research, University of New Brunswick.

Willoughby, T. and **E. Wood** (1994), "Elaborative interrogation examined at encoding and retrieval", *Learning and Instruction,* Vol. 4, Issue 2, Elsevier Ltd., pp. 139-149.

Winne, P.H. (2001), "Self-regulated learning viewed from models of information processing", in B.J. Zimmerman and D.H. Schunk (eds.), *Self-regulated learning and academic achievement: theoretical perspectives,* Lawrence Erlbaum Associates, Inc., Mahwah, N.J., pp. 153-189.

Zimmerman, B.J. (1999), "Commentary: toward a cyclically interactive view of self-regulated learning", *International Journal of Educational Research,* Vol. 31, Issue 6, Elsevier Ltd., pp.545-551.

Zimmerman, B. J. and **M. Martinez-Pons** (1990), "Student differences in self-regulated learning: Relating grade, sex and giftedness to self efficacy and strategy use", *Journal of Educational Psychology,* Vol. 82, No. 1, American Psychological Association, Washington, D.C., pp. 51-59.