2

Science performance among 15-year-olds

This chapter defines the notion of science literacy and how it is measured in PISA 2015. It also shows how close countries are to equipping all their students with a baseline level of proficiency in science. This would mean that, when students leave compulsory education, they are at least able to provide possible explanations for scientific phenomena in familiar contexts and to draw appropriate conclusions from data derived from simple investigations. The chapter also discusses the extent to which young adults have acquired a scientific mindset – that is, positive dispositions towards scientific methods of enquiry and towards discussion of science-related topics.

A note regarding Israel

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

An understanding of science, and of science-based technology, is necessary not only for those whose careers depend on it directly, but also for any citizen who wishes to make informed decisions related to the many controversial issues under debate today – from more personal issues, such as maintaining a healthy diet, to local issues, such as how to manage waste in big cities, to more global and far-reaching issues, such as the costs and benefits of genetically modified crops or how to prevent and mitigate the catastrophic consequences of global warming.

Science education in primary and secondary school should ensure that by the time students leave school they can understand and engage in discussions about the science and technology-related issues that shape our world. Most current curricula for science education are designed on the premise that an understanding of science is so important that the subject should be a central feature in every young person's education (OECD, 2016b).

What the data tell us

- Singapore outperforms all other participating countries/economies in science. Japan, Estonia, Finland and Canada, in descending order of mean performance, are the four highest-performing OECD countries.
- Some 7.7% of students across OECD countries are top performers in science, meaning that they are proficient at Level 5 or 6. About one in four (24.2%) students in Singapore, and more than one in seven students in Chinese Taipei (15.4%), Japan (15.3%) and Finland (14.3%) perform at this level.
- Mean performance in science improved significantly between 2006 and 2015 in Colombia, Israel, Macao (China), Portugal, Qatar and Romania. Over this period, Macao (China), Portugal and Qatar reduced the share of low-achieving students performing below Level 2, and simultaneously increased the share of students performing at or above Level 5.
- In 33 countries and economies, the share of top performers in science is larger among boys than among girls. Finland is the only country in which girls are more likely to be top performers than boys. At the same time, in most countries, boys and girls are equally able to complete the easiest science tasks in the PISA test.
- Students who score low in science are less likely to agree that scientific knowledge is tentative and to believe that scientific approaches to enquiry, such as repeating experiments, are a good way to acquire new knowledge.

HOW PISA DEFINES SCIENCE LITERACY

PISA 2015 focused on science as the major domain, and defines science literacy as "the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen". A scientifically literate person is willing to engage in reasoned discourse about science and technology. This requires the competencies to explain phenomena scientifically, to evaluate and design scientific enquiry, and to interpret data and evidence scientifically (for a detailed description of science literacy, see the *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematics and Financial Literacy*, OECD, 2016b).

Performance in science requires three forms of knowledge: content knowledge, knowledge of the standard methodological procedures used in science, and knowledge of the reasons and ideas used by scientists to justify their claims. Explaining scientific and technological phenomena, for instance, demands knowledge of the content of science. Evaluating scientific enquiry and interpreting evidence scientifically also require an understanding of how scientific knowledge is established and the degree of confidence with which it is held.

The definition of science literacy recognises that there is an affective element to a student's competency: students' attitudes or dispositions towards science can influence their level of interest, sustain their engagement and motivate them to take action (Osborne, Simon and Collins, 2003; Schibeci, 1984).

The use of the term "science literacy" underscores PISA's aim not only to assess what students know in science, but also what they can do with what they know, and how they can creatively apply scientific knowledge to real-life situations. In the remaining parts of this chapter, "science" is also used to refer to the "science literacy" measured in PISA.

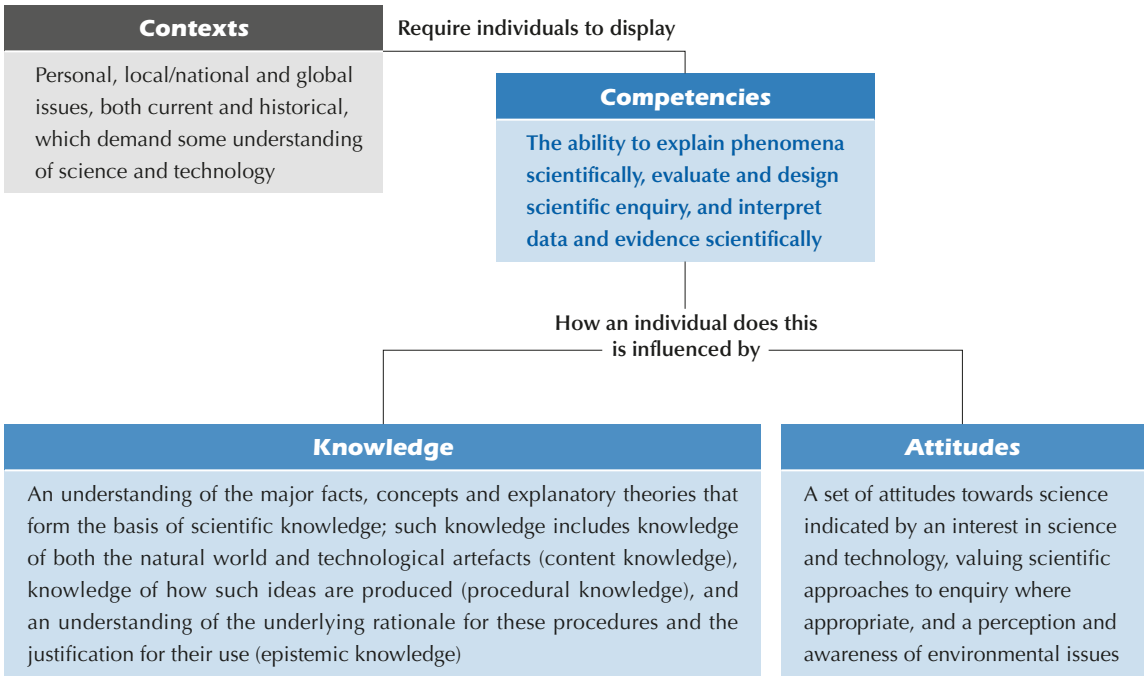
Described in this way, literacy in science is not an attribute that a student has or does not have; rather, it can be acquired to a greater or lesser extent, and is influenced both by knowledge of and about science, and by attitudes towards science.

The concept of science literacy in PISA refers to a knowledge of both *science* and *science-based technology*, even though science and technology do differ in their purposes, processes and products. Technology seeks the optimal solution to a human problem, and there may be more than one optimal solution. In contrast, science seeks the answer to a specific question about the natural, material world. Nevertheless, the two are closely related, and science-literate individuals are expected to be able and willing to engage in reasoned discourse, and make informed decisions, about both science and technology. For instance, individuals make decisions and choices that influence the directions of new technologies (such as the decision to drive a smaller, more fuel-efficient car). Scientifically literate individuals are expected to make more informed choices. They should also be able to recognise that, while science and technology are often a source of solutions, paradoxically, they can also be seen as a source of risk, generating new problems that can only be solved through the use of science and technology.

The PISA 2015 framework for assessing science literacy

Figure I.2.1 presents presents an overview of the main aspects of the PISA 2015 framework for science that was established and agreed by the countries and economies participating in PISA, and how the aspects are related to each other. The central box, highlighted in blue, lists the three competencies that lie at the heart of the PISA definition of science literacy: explaining phenomena scientifically, evaluating and designing scientific enquiry, and interpreting data and evidence scientifically. Students use these competencies in specific contexts that demand some understanding of science and technology; these contexts generally relate to local or global issues. Students' ability to apply their competencies to a specific science context is influenced by both their attitudes towards science, scientific methods and the underlying issue, and by their knowledge of science ideas and how they are produced and justified.

Figure I.2.1 ■ **Aspects of the science assessment framework for PISA 2015**



The PISA 2015 framework for assessing science in PISA builds on the previous framework, developed for the 2006 assessment. The major difference is that the notion of "knowledge about science", which was referred to in the PISA 2006 definition as an "understanding of the characteristic features of science as a form of human knowledge and enquiry", has been defined more clearly and split into two components – procedural knowledge and epistemic knowledge (i.e. knowledge of the nature and origin of scientific understanding). Several changes in the test design, most notably the move from paper-based to computer-based delivery, also influenced the development of the assessment tasks, as is explained in greater detail below.

Each of the tasks used for the assessment of students' performance in science has been mapped against the different aspects of the framework, as well as against two additional dimensions (response format and cognitive demand), in order to create a balanced assessment that covers the full framework. The distribution of items across framework categories reflects a consensus view among the experts consulted on the relative weight of these components in the definition of science literacy (OECD, 2016b). The six dimensions used to classify items are explained in detail below and are summarised in Figure I.2.2. Three of the six – scientific competencies, knowledge types and content areas – are reporting categories: for each of them, it is possible to contrast student performance in the various subcategories by using subscales.

Figure I.2.2 ■ **Categories describing the items constructed for the PISA 2015 science assessment**

Reporting categories			Further categories to ensure a balanced assessment		
Scientific competencies	**Knowledge types**	**Content areas**	**Response types**	**Cognitive demand**	**Contexts**
Explain phenomena scientifically	Content	Physical systems	Simple multiple choice	Low	Personal
Evaluate and design scientific enquiry	Procedural[1]	Living systems	Complex multiple choice	Medium	Local/National
Interpret data and evidence scientifically	Epistemic[1]	Earth and space systems	Constructed response	High	Global

1. While distinct from a theoretical point of view, the procedural and epistemic knowledge categories form a single reporting category.

Scientific competencies

According to the PISA definition, a science-literate person is able and willing to engage in reasoned discourse about science and technology. This requires the competencies to:

- **Explain phenomena scientifically** – recognise, offer and evaluate explanations for a range of natural and technological phenomena.
- **Evaluate and design scientific enquiry** – describe and appraise scientific investigations and propose ways of addressing questions scientifically.
- **Interpret data and evidence scientifically** – analyse and evaluate data, claims and arguments in a variety of representations and draw appropriate scientific conclusions.

That the three science competencies are central to the definition of science literacy reflects a view that science is best seen as an ensemble of practices for generating, evaluating and discussing knowledge that is common across all of the natural sciences. Fluency with these practices reflects greater competency, and distinguishes the expert scientist from the novice. While it would be unreasonable to expect a 15-year-old student to have the expertise of a professional scientist, a scientifically literate student can be expected to appreciate the role and significance of these practices and demonstrate a basic proficiency in them.

The competency "explain phenomena scientifically", defined as the ability to recognise, offer and evaluate explanations for a range of natural and technological phenomena, is evident when students recall and apply appropriate scientific knowledge; identify, use and generate explanatory models and representations; make and justify appropriate predictions; offer explanatory hypotheses; and explain the potential implications of scientific knowledge for society.

The competency "evaluate and design scientific enquiry" is required to evaluate reports of scientific findings and investigations critically. It is defined as the ability to describe and appraise scientific investigations and propose ways of addressing questions scientifically. It is reflected in the behaviour of students who identify the question explored in a given scientific study; distinguish questions that can be investigated scientifically from those that cannot; propose a way of exploring a given question scientifically; evaluate ways of exploring a given question scientifically; and describe and evaluate how scientists ensure the reliability of data, and the objectivity and generalisability of explanations.

The competency "interpret data and evidence scientifically" is defined as the ability to analyse and evaluate scientific data, claims and arguments in a variety of representations, and draw appropriate conclusions. Students who can interpret data and evidence scientifically can transform data from one representation to another; analyse and interpret data and draw appropriate conclusions; identify the assumptions, evidence and reasoning behind science-related texts; distinguish between arguments that are based on scientific evidence and theory and those based on other considerations; and contrast and evaluate scientific arguments and evidence from different sources.

The 184 science-related test items – the equivalent of around six hours of test material – from which the PISA 2015 assessment of science was assembled can be classified into categories related to these three competencies according to the main demand of the task. Among all science-related items, 48% (89 items, or the equivalent of almost three hours) mainly draw on students' ability to explain phenomena scientifically, 21% (39 items, or slightly more than one hour) on the ability to evaluate and design scientific enquiry, and 30% (56 items, or almost two hours) on the ability to interpret data and evidence scientifically (see Annex C2).

Knowledge categories

Each of the scientific competencies requires some content knowledge (knowledge of theories, explanatory ideas, information and facts), but also an understanding of how such knowledge has been derived (procedural knowledge) and of the nature of that knowledge (epistemic knowledge).

"Procedural knowledge" refers to knowledge about the concepts and procedures that are essential for scientific enquiry, and that underpin the collection, analysis and interpretation of scientific data. In the quest to explain phenomena in the material world, science proceeds by testing hypotheses through empirical enquiry. Empirical enquiry relies on certain standard procedures to obtain valid and reliable data. Students are expected to know these procedures and related concepts, such as: the notion of dependent and independent variables; the distinction between different types of measurement (qualitative and quantitative, categorical and continuous); ways of assessing and minimising uncertainty (such as repeating measurements); the strategy of controlling variables and its role in experimental design; and common ways of presenting data. It is expected, for instance, that students will know that scientific knowledge is associated with differing degrees of certainty, depending on the nature and quantity of empirical evidence that has accumulated over time.

"Epistemic knowledge" refers to an understanding of the nature and origin of knowledge in science, and reflects students' capacity to think and engage in reasoned discourse as scientists do. Epistemic knowledge is required to understand the distinction between observations, facts, hypotheses, models and theories, but also to understand why certain procedures, such as experiments, are central to establishing knowledge in science.

Slightly over half of all the science-related items in PISA 2015 (98 out of 184) require mainly content knowledge, 60 require procedural knowledge, and 26 require epistemic knowledge.

Content areas

Knowledge can also be classified according to the major scientific fields to which it pertains. Fifteen-year-old students are expected to understand major explanatory ideas and theories from the fields of physics, chemistry, biology, earth and space sciences, and how they apply in contexts where the elements of knowledge are interdependent or interdisciplinary. Items used in the assessment are classified into three content areas: physical systems, living systems, and earth and space systems.[1] Examples of knowledge that 15-year-olds are expected to have acquired include an understanding of the particle model of matter (physical systems), the theory of evolution by natural selection (living systems), and the history and scale of the universe (earth and space systems). About one-third of all the science-related items in PISA 2015 (61 out of 184) relate to physical systems, 74 to living systems, and the remaining 49 to earth and space systems.

Context of assessment items

The real-world issues used as stimuli and items for the assessment of science literacy in 2015 can also be classified by the context in which they are set. Three context categories identify the broad areas of life in which the test problems may arise: "personal", which are contexts related to students' and families' daily lives; "local/national", which are contexts related to the community in which students live; and "global", which are contexts defined by life across the world. An item relating to a fossil fuel issue, for instance, may be classified as personal if it explores energy-saving behaviours, as local/national if it addresses the environmental impact on air quality, and as global, if it examines the link between fossil fuel consumption and the concentration of carbon dioxide in the atmosphere.

The PISA 2015 science assessment is not an assessment of specific contexts; rather, the contexts are used to elicit specific science-related tasks. Therefore, a broad range of personal, local/national and global contexts was included in the assessment.

Attitudes

Peoples' attitudes and beliefs play a significant role in their interest, attention and response to science and technology. The PISA definition of science literacy recognises that a student's response to a science-related issue requires more

than skills and knowledge; it also depends on how able and "willing" the student is "to engage" with the issue. In PISA 2015, students' attitudes, beliefs and values were examined through students' responses to questions in the student questionnaire rather than through their performance on test items. A major distinction among science-related attitudes is between attitudes towards science (e.g. interest in different content areas of science) and scientific attitudes. The former set of attitudes is examined in greater detail in the next chapter. Students' beliefs about science knowledge and knowing (epistemic beliefs), which indicate whether students value scientific approaches to enquiry and are part of the latter set of attitudes, are analysed at the end of this chapter.

Computer-based assessment of science

Computer delivery of the PISA 2015 assessment has made it possible to expand what the PISA science test can assess, compared to previous paper-based versions of PISA tests. For instance, PISA 2015 for the first time assessed students' ability to conduct scientific enquiry by asking them to design (simulated) experiments and interpret the resulting evidence. This was made possible through the use of interactive presentations, where students' actions determined what they saw on the screen. Twenty-four items included in the main study (or about 13%) were interactive, but they were kept confidential so that they can be used in future assessments to measure trends.

The PISA 2015 field-trial unit *RUNNING IN HOT WEATHER*, available online at www.oecd.org/pisa and described in Annex C1, provides an illustration of how interactive science items work. It asks students to collect data on the water loss and body temperature of a runner after a one-hour run under different temperature and humidity conditions. After moving sliders that appear on the screen to the desired temperature and humidity levels, students can run one or more simulations whose results are recorded on the screen and must be used in order to answer the questions in that unit.

Questions based on interactive presentations can focus on the ability to interpret data and evidence scientifically (e.g. Question 1 in *RUNNING IN HOT WEATHER*), on the ability to explain phenomena scientifically (e.g. Question 2), or on the ability to evaluate and design scientific enquiry (e.g. Question 3), and can relate to all content areas and types of knowledge. The relative difficulty or complexity of a particular question was not related to whether the item was presented as interactive or static.

Computer delivery of test items also allowed for a greater variety of contexts to be included in the assessment, and to convey situations of motion and change (e.g. chemical reactions) in a more realistic and motivating way, through the use of animations.

Response types used in the assessment of science

Three broad categories of response formats were used in the PISA 2015 science assessment: simple multiple choice, complex multiple choice, and constructed response. Within each category, new response formats, in addition to those that were also used in paper-based tests, were used in the computer-based science assessment. About one-third of the items can be classified in each category:

- simple multiple choice: items calling for
  - selection of a single response from four options
  - selection of a "hot spot", an answer that is a selectable element within a graphic or text
- complex multiple choice: items calling for
  - responses to a series of related "Yes/No" questions that are scored as a single item (the typical format in 2006)
  - selection of more than one response from a list
  - completion of a sentence by selecting choices from a drop-down menu to fill multiple blanks
  - "drag-and-drop" responses, allowing students to move elements on screen to complete a task of matching, ordering or categorising
- constructed response: items calling for written or drawn responses. Constructed-response items in science typically call for a written response ranging from a phrase to a short paragraph (e.g. two to four sentences of explanation). A small number of constructed-response items call for a drawing (e.g. a graph or diagram). In a computer-based assessment, any such item is supported by simple drawing applications that are specific to the response required. In general, these items cannot be machine scored; they require the professional judgement of trained coders to assign the responses to defined categories. To ensure that the response-coding process yields reliable and cross-nationally comparable results, detailed guidelines and training were provided. All of the procedures to ensure consistency of coding within and between countries are detailed in the *PISA 2015 Technical Report* (OECD, forthcoming).

Cognitive demand of items

A novel feature of the PISA 2015 science assessment was the explicit attempt to cover different levels of cognitive demand across all three types of science competencies and knowledge. Cognitive demand, sometimes referred to as "depth of knowledge", refers to the type of mental processes required to complete an item. In large part, it determines an item's level of difficulty, more than the response format or a student's familiarity with the underlying science content.

The cognitive demand – and thus difficulty – of items is influenced by four factors:

- the number and degree of complexity of the elements of knowledge in the item
- students' level of familiarity with and prior knowledge of the content, procedural and epistemic knowledge involved
- the cognitive operation required by the item, e.g. recall, analysis and/or evaluation
- the extent to which forming a response depends on models or abstract scientific ideas.

To ensure a balanced assessment of science, three levels of cognitive demand are identified:

- **Low depth of knowledge:** Items requiring the student to carry out a one-step procedure, such as recalling a single fact, term, principle or concept, or locating a single point of information from a graph or table.
- **Medium depth of knowledge:** Items requiring the student to use and apply conceptual knowledge to describe or explain phenomena, select appropriate procedures involving two or more steps, organise/display data, or interpret and use simple data sets and graphs.
- **High depth of knowledge:** Items requiring students to analyse complex information or data, synthesise or evaluate evidence, justify claims, reason (given various sources), or develop a plan with which to approach a problem.

Of the 184 items included in the PISA 2015 science assessment, 56 (or about 30%) are classified in the "low depth of knowledge" category, 15 (or about 8%) in the "high depth of knowledge" category, and the majority (113 items, or 61%) in the "medium" category.

Examples of items representing the different categories

Figure I.2.3 summarises how the sample items from the PISA 2015 main study (described in greater detail in Annex C1 and available on line at www.oecd.org/pisa) are categorised.

Figure I.2.3 ■ **Classification of sample items**

By competency, knowledge and content categories, depth of knowledge, response type and context

Item/Question	Scientific competency	Knowledge type	Content area	Cognitive demand	Response type	Context
SUSTAINABLE FISH FARMING, Question 1	Explain phenomena scientifically	Content	Living	Medium	Complex multiple choice	Local/ National
SUSTAINABLE FISH FARMING, Question 2	Interpret data and evidence scientifically	Content	Living	Low	Simple multiple choice	Local/ National
SUSTAINABLE FISH FARMING, Question 3	Explain phenomena scientifically	Content	Physical	Low	Simple multiple choice	Local/ National
SLOPE-FACE INVESTIGATION, Question 1	Evaluate and design scientific enquiry	Epistemic	Earth and space	Medium	Constructed response	Local/ National
SLOPE-FACE INVESTIGATION, Question 3	Interpret data and evidence scientifically	Epistemic	Earth and space	High	Constructed response	Local/ National
METEOROIDS AND CRATERS, Question 1	Explain phenomena scientifically	Content	Physical	Low	Simple multiple choice	Global
METEOROIDS AND CRATERS, Question 2	Explain phenomena scientifically	Content	Earth and space	Low	Complex multiple choice	Global
METEOROIDS AND CRATERS, Question 3A	Explain phenomena scientifically	Content	Earth and space	Low	Complex multiple choice	Global
METEOROIDS AND CRATERS, Question 3B	Explain phenomena scientifically	Content	Earth and space	Medium	Complex multiple choice	Global
BIRD MIGRATION, Question 1	Explain phenomena scientifically	Content	Living	Medium	Simple multiple choice	Global
BIRD MIGRATION, Question 2	Evaluate and design scientific enquiry	Procedural	Living	High	Constructed response	Global
BIRD MIGRATION, Question 3	Interpret data and evidence scientifically	Procedural	Living	Medium	Complex multiple choice	Global

HOW THE PISA 2015 SCIENCE RESULTS ARE REPORTED

In 57 countries/economies, including all OECD countries, the PISA 2015 test was conducted on computers. The paper-based form was used in 15 countries/economies as well as in Puerto Rico, an unincorporated territory of the United States. The countries/economies that administered the paper-based test in 2015 are: Albania, Algeria, Argentina, the Former Yugoslav Republic of Macedonia (hereafter "FYROM"), Georgia, Indonesia, Jordan, Kazakhstan, Kosovo, Lebanon, Malta, Moldova, Romania, Trinidad and Tobago, and Viet Nam. Only the computer-based test fully covers the new aspects of the science framework for PISA 2015. The paper-based test used only items developed in previous cycles, which represent about half of all the items used in the computer-based assessments. Nevertheless, the procedures used to develop the tests and to analyse and scale student responses were the same for both sets of countries/economies that participated in PISA 2015. And while the science test is not equivalent across the two modes of delivery, results of the paper-based and computer-based tests in 2015 are linked through common items. The results of both are reported on the same scale as the results of previous assessments, so that all countries can be directly compared across modes and across time (see Box I.2.3).[2]

How the PISA 2015 science test was designed, analysed and scaled

This section summarises the test development and scaling procedures used to ensure that results of the PISA 2015 test are comparable across countries and with the results of previous PISA assessments. These procedures are described in greater detail in the *PISA 2015 Technical Report* (OECD, forthcoming). While the development and selection of test questions mostly followed procedures established in previous PISA cycles, several changes were introduced in the administration procedures (including the move from paper- to computer-based delivery and an improved design of test forms) and in the scaling procedures. The impact of these changes on comparing student performance over time is further discussed in Box I.2.3 and Annex A5.

How test questions were developed and selected

The test material had to meet several requirements:

- Test items had to meet the requirements and specifications of the framework for PISA 2015 that was established and agreed upon by the participating countries and economies. The content, cognitive demands and contexts of the items had to be deemed appropriate for a test for 15-year-olds.
- Items had to be of curricular relevance for 15-year-olds in participating countries and economies and appropriate in the respective cultural contexts. It is inevitable that not all tasks in the PISA assessment are equally appropriate in different cultural contexts and equally relevant in different curricular and instructional contexts. But PISA asked experts from every participating country to identify those tasks from the PISA tests that they considered most appropriate for an international test, and these ratings were considered when selecting items for the assessment.
- Items had to meet stringent standards of technical quality and international comparability. In particular, the professional translation and verification of items and an extensive field trial ensured the linguistic equivalence of test questions across the more than 70 languages in which PISA 2015 was conducted. The field trial also served to verify the psychometric equivalence of the instruments, which was further examined before scaling the results of the main study (see Annex A5).
- A sufficient number of items from previous assessments had to be included in order to allow for comparisons with previous rounds of PISA and to continue measuring trends.

Items for the science assessment were selected from a pool of diverse material with a broad range of authors from different cultures and countries.

Just under 50% of the PISA 2015 science items were initially developed for delivery on paper in the PISA 2006 assessment of science and have been kept strictly confidential thereafter. These "trend units" provide the basis for measuring changes in student performance over time, and for linking the PISA 2015 science scale to the existing PISA science scale. All trend items used in PISA 2015 had to be adapted for delivery on computer (also see *PISA 2015 Technical Report* [OECD, forthcoming], Chapter 2). The equivalence between the paper- and computer-based versions of trend items used to measure student proficiency in science, reading and mathematics was assessed on a diverse population of students from all countries that participated in PISA 2015 as part of an extensive field trial. The results of this mode study informed the selection of items and the scaling of student responses for the PISA 2015 main survey (see Box I.2.3).

Slightly more than half of the items used in the assessment were newly developed for computer delivery in PISA 2015. Authors in 14 countries, with contributions from national teams, members of the PISA science expert group, and the PISA International Consortium, created stimulus material and questions that reflect the content, contexts and approaches relevant to students in a large number of PISA-participating countries and economies. Experts reviewed wording and other features of the items, then the items were tested among classes of 15-year-old students in the field trial.

The items were extensively field tested in all countries and economies that participated in the PISA 2015 assessment. Local science experts in each participating country and economy provided detailed feedback on the curricular relevance, appropriateness and potential interest for 15-year-olds. At each stage, material was considered for rejecting, revising or keeping in the pool of potential items. Finally, the international science expert group formulated recommendations as to which items should be included in the main survey instruments. The final set of items selected for the main survey was also subject to reviews by all countries and economies. During those reviews, countries/ economies provided recommendations in relation to: item suitability for assessing the competencies enumerated in the framework; the items' acceptability and appropriateness at the national level; and the overall quality of the assessment instruments, to ensure they were of the highest standard possible. This selection was balanced across the various categories specified in the science framework and spanned a range of levels of difficulty, so that the entire pool of items could measure performance across all science competencies and knowledge types, and across a broad range of content areas and student abilities (for further details, see the *PISA 2015 Technical Report* [OECD, forthcoming]).

Test items were generally developed within "units" that included some stimulus material and one or more questions related to the stimulus.

Altogether, the 184 items that were developed and selected for the PISA 2015 science assessment represent the equivalent of six hours of test questions. Of these items, 85 questions (the equivalent of about three hours) are trend tasks, which were used in previous PISA surveys, and 99 questions (another three hours) are new science tasks. Trend tasks that had originally been developed for paper-based assessments were adapted for computer-based delivery in 57 countries/ economies. They were included in their original paper-based form in the countries/economies that conducted the PISA 2015 test with paper and pencil. New tasks were developed for computer-based delivery and were only included in the tests in the 57 countries that conducted the PISA 2015 test on computer.

How the test forms were designed

In order to ensure that the assessment covered a wide range of content, with the understanding that each student could complete only a limited set of tasks, the full set of tasks was distributed across a range of test forms with overlapping content. Each student thus completed only a fraction of all items, depending on which test form was randomly assigned to him or her. All forms contained an hour-long sequence of science questions, and therefore all students completed about one hour of testing in science – or about 30 items.

Half of the students sat the science test during the first hour of the assessment, and half sat the test during the second hour, after a short break. During the other hour of testing, students worked on sequences of tasks from either one or two of the following domains: reading, mathematics, and in 50 countries and economies, collaborative problem solving, so that all students completed two hours of testing in two or three domains, including science. In 15 countries and economies, a subset of the students in the PISA sample also completed a test of financial literacy after completing the main PISA test and questionnaire. The number and sequence of test domains and of tasks depended on the test form, which was assigned to students by a random draw.

How student responses were analysed and scaled

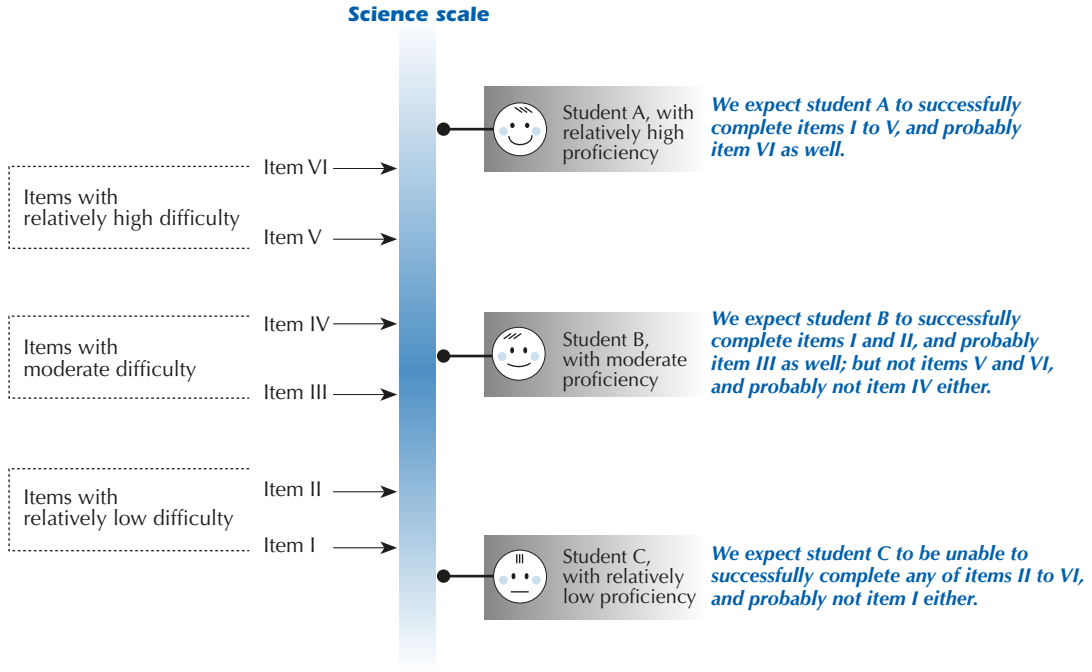
While different students saw different questions, the test design, which was built on those used in previous PISA assessments, made it possible to construct a continuous scale of proficiency in science, so that each test-taker's performance is associated with a particular point on the scale that indicates his or her estimated science proficiency, and the likelihood that he or she responds correctly to a particular question (higher values on the scale indicate greater proficiency). A description of the modelling technique used to construct this scale can be found in the *PISA 2015 Technical Report* (OECD, forthcoming).

The relative difficulty of tasks was estimated by determining the proportion of test-takers who answer each question correctly. Task difficulty is reported on the same scale as student proficiency (higher values correspond, in this case, to more difficult items). In PISA, the difficulty of a task is defined as the point on the scale where there is at least a 62% probability of a correct response by students who score at or above that point.[3] A single continuous scale shows

the relationship between the difficulty of questions and the proficiency of test-takers (Figure I.2.4). By constructing a scale that shows the difficulty of each question, it is possible to locate the level of science literacy that the question demands. By showing the proficiency of each test-taker on the same scale, it is possible to describe each test-taker's level of science literacy.

Just as the sample of students who sat the PISA test in 2015 was drawn to represent all 15-year-old students in the participating countries and economies, so the individual test questions used in the assessment were designed to represent the definition of literacy in science described above. Estimates of student proficiency reflect the kinds of tasks students would be expected to perform successfully. This means that students are likely to be able to successfully answer questions located at or below the difficulty level associated with their own position on the scale. Conversely, they are unlikely to be able to successfully answer questions above the difficulty level associated with their position on the scale.

Figure I.2.4 ▪ **Relationship between questions and student performance on a scale**



The higher a student's proficiency level is located above a given test question, the more likely is he or she to answer the question (and other questions of similar difficulty) successfully. The further the student's proficiency is located below a given question, the less likely is he or she to be able to answer the question (and other questions of similar difficulty) successfully.

Reporting scales for PISA 2015

PISA 2015 provides an overall science scale, which draws on all of the science questions in the assessment, as well as (for countries/economies that used the full set of PISA 2015 science items, i.e. those that administered the PISA 2015 test on computers) scales for the three science competencies, the three content areas and two of the broad knowledge-type categories defined earlier in this chapter. (A single scale for both procedural and epistemic knowledge was constructed because there were too few epistemic knowledge items to support the construction of a continuous scale of epistemic knowledge with desirable properties.)[4] The metric for the overall science scale is based on a mean for OECD countries of 500 points and a standard deviation of 100 points that were set in PISA 2006 when the PISA science scale was first developed.[5] The items that were common to both the 2006 and 2015 test instruments, and were found to measure science competencies comparably in the paper- and computer-based modes, allow for a link to be made with the earlier scale. Annex A5 describes how the PISA 2015 scale was equated to the PISA 2006 scale.

How science proficiency levels are defined in PISA 2015

To help users interpret what student scores mean in substantive terms, PISA scales are divided into proficiency levels. For PISA 2015, the range of difficulty of science tasks is represented by seven levels of science proficiency: six levels that are aligned with the levels used in describing the outcomes of PISA 2006 (ranging from the highest, Level 6, to Level 1a, formerly known as Level 1). At the bottom of the scale, a new Level 1b is described, based on some of the easiest tasks included in the assessment, to indicate the knowledge and skills of some of the students performing below Level 1a (in previous PISA reports, these students were included among those scoring "below Level 1").

Based on the cognitive demands of tasks that are located within each level, descriptions of each of these levels have been generated to define the kinds of knowledge and skills needed to complete those tasks successfully. Individuals with proficiency within the range of Level 1b are likely to be able to complete Level 1b tasks, but are unlikely to be able to complete tasks at higher levels. Level 6 includes tasks that pose the greatest challenge in terms of the depth of science knowledge and competencies needed to complete them successfully. Students with scores in this range are likely to be able to complete tasks located at this level, as well as all the other PISA science tasks (see the following section for a detailed description of the proficiency levels in science).

Figure I.2.5 shows the location on the science scale of some of the items used in the PISA 2015 assessment of science. These items are only a small sample of all the items used in the assessment, and are presented in greater detail in Annex C1 and on line at www.oecd.org/pisa. While no item at Level 1a and at Level 5 are included among the released main survey items shown in the figure, there were 10 items at Level 1a among the 184 science items used in PISA 2015, and 20 items at Level 5. Since PISA is a recurring assessment, it is useful to retain a sufficient number of questions over successive PISA assessments in order to generate trend data over time.

Figure I.2.5 ▪ **Map of selected science questions illustrating proficiency levels**

Level	Lower score limit	Question	Question difficulty (in PISA score points)
6	708	**SUSTAINABLE FISH FARMING** – Question 1 (S601Q01)	740
5	633		
4	559	**BIRD MIGRATION** – Question 2 (S656Q02)	630
		SLOPE–FACE INVESTIGATION – Question 3 (S637Q05)	589
		SUSTAINABLE FISH FARMING – Question 3 (S601Q04)	585
		BIRD MIGRATION – Question 3 (S656Q04)	574
3	484	**SLOPE–FACE INVESTIGATION** – Question 1 (S637Q01)	517
		BIRD MIGRATION – Question 1 (S656Q01)	501
2	410	**METEOROIDS AND CRATERS** – Question 1 (S641Q01)	483
		SUSTAINABLE FISH FARMING – Question 2 (S601Q02)	456
		METEOROIDS AND CRATERS – Question 2 (S641Q02)	450
		METEOROIDS AND CRATERS – Question 3B (S641Q04)	438
1a	335		
1b	261	**METEOROIDS AND CRATERS** – Question 3A (S641Q03)	299

For all levels, the descriptions have been updated to reflect the new categories in the PISA 2015 framework and the large number of new items developed for PISA 2015. Strictly speaking, the updated descriptions only apply to countries that conducted the PISA 2015 test on computer. While the results of the paper-based test conducted in 15 countries/economies can be reported on the same scale as the results of the computer-based test, these countries only used items that were originally developed in PISA 2006.

Figure I.2.6 provides descriptions of the science competencies, knowledge and understanding required at each level of the science literacy scale, and the average proportion of students across OECD countries who perform at each of these proficiency levels.

Figure I.2.6 ■ **Summary description of the seven levels of proficiency in science in PISA 2015**

Level	Lower score limit	Characteristics of tasks
6	708	At Level 6, students can draw on a range of interrelated scientific ideas and concepts from the physical, life and earth and space sciences and use content, procedural and epistemic knowledge in order to offer explanatory hypotheses of novel scientific phenomena, events and processes or to make predictions. In interpreting data and evidence, they are able to discriminate between relevant and irrelevant information and can draw on knowledge external to the normal school curriculum. They can distinguish between arguments that are based on scientific evidence and theory and those based on other considerations. Level 6 students can evaluate competing designs of complex experiments, field studies or simulations and justify their choices.
5	633	At Level 5, students can use abstract scientific ideas or concepts to explain unfamiliar and more complex phenomena, events and processes involving multiple causal links. They are able to apply more sophisticated epistemic knowledge to evaluate alternative experimental designs and justify their choices and use theoretical knowledge to interpret information or make predictions. Level 5 students can evaluate ways of exploring a given question scientifically and identify limitations in interpretations of data sets including sources and the effects of uncertainty in scientific data.
4	559	At Level 4, students can use more complex or more abstract content knowledge, which is either provided or recalled, to construct explanations of more complex or less familiar events and processes. They can conduct experiments involving two or more independent variables in a constrained context. They are able to justify an experimental design, drawing on elements of procedural and epistemic knowledge. Level 4 students can interpret data drawn from a moderately complex data set or less familiar context, draw appropriate conclusions that go beyond the data and provide justifications for their choices.
3	484	At Level 3, students can draw upon moderately complex content knowledge to identify or construct explanations of familiar phenomena. In less familiar or more complex situations, they can construct explanations with relevant cueing or support. They can draw on elements of procedural or epistemic knowledge to carry out a simple experiment in a constrained context. Level 3 students are able to distinguish between scientific and non-scientific issues and identify the evidence supporting a scientific claim.
2	410	At Level 2, students are able to draw on everyday content knowledge and basic procedural knowledge to identify an appropriate scientific explanation, interpret data, and identify the question being addressed in a simple experimental design. They can use basic or everyday scientific knowledge to identify a valid conclusion from a simple data set. Level 2 students demonstrate basic epistemic knowledge by being able to identify questions that can be investigated scientifically.
1a	335	At Level 1a, students are able to use basic or everyday content and procedural knowledge to recognise or identify explanations of simple scientific phenomenon. With support, they can undertake structured scientific enquiries with no more than two variables. They are able to identify simple causal or correlational relationships and interpret graphical and visual data that require a low level of cognitive demand. Level 1a students can select the best scientific explanation for given data in familiar personal, local and global contexts.
1b	261	At Level 1b, students can use basic or everyday scientific knowledge to recognise aspects of familiar or simple phenomenon. They are able to identify simple patterns in data, recognise basic scientific terms and follow explicit instructions to carry out a scientific procedure.

A CONTEXT FOR COMPARING THE SCIENCE PERFORMANCE OF COUNTRIES AND ECONOMIES

Comparing science performance, and performance in school more generally, poses numerous challenges. When teachers give a science test in a classroom, students with varying abilities, attitudes and social backgrounds are required to respond to the same set of tasks. When educators compare the performance of schools, the same test is used across schools that may differ significantly in the structure and sequencing of their curricula, in the pedagogical emphases and instructional methods applied, and in the demographic and social contexts of their student populations. Comparing the performance of education systems across countries adds more layers of complexity, because students are given tests in different languages, and because the social, economic and cultural context of the countries that are being compared are often very different.

However, while students within a country may learn in different contexts according to their home background and the school they attend, their performance is measured against common standards. For example, when they become adults, they will all face common challenges and will often have to compete for the same jobs. Similarly, in a global economy, the benchmark for success in education is no longer improvement by national standards alone, but increasingly, in relation to the best-performing education systems around the world. As difficult as international comparisons are, they are important for educators, and PISA goes to considerable lengths to ensure that such comparisons are valid and fair.

This section discusses countries' science performance in the context of important economic, demographic and social factors that can influence assessment results. It provides a context for interpreting the results that are presented later in the chapter.

PISA's stringent standards for sampling limit the possible exclusion of students and schools and the impact of non-response. These standards are applied to ensure that, for all adjudicated countries, economies and subnational regions, the results support conclusions that are valid for the PISA target population (all students between 15 years and 3 [completed] months and 16 years and 2 [completed] months at the beginning of the testing period, attending educational institutions located within the adjudicated entity, and in grade 7 or higher).

But when interpreting PISA results with regard to the overall population of 15-year-olds, sample coverage must be assessed with respect to this wider population. In most OECD countries and in many partner countries and economies, the target population represents more than 80% of the estimated number of 15-year-olds in the country, so that results can be extended, with some caution but with a high degree of confidence, beyond the PISA target population to all 15-year-olds. By contrast, in a few countries participating in PISA, including OECD countries Mexico and Turkey, the share of out-of-school 15-year-olds, or the number of 15-year-olds who are still in primary education (in grade 6 or lower), represents a significant fraction of the PISA age cohort. "Coverage index 3", discussed in Chapter 6, provides an estimate of the share of the age cohort covered by PISA. It varies from 49% in Viet Nam to more than 95% in Finland, Germany, Ireland, Malta, the Netherlands, the Russian Federation (hereafter "Russia"), Singapore and Switzerland (Table I.6.1).

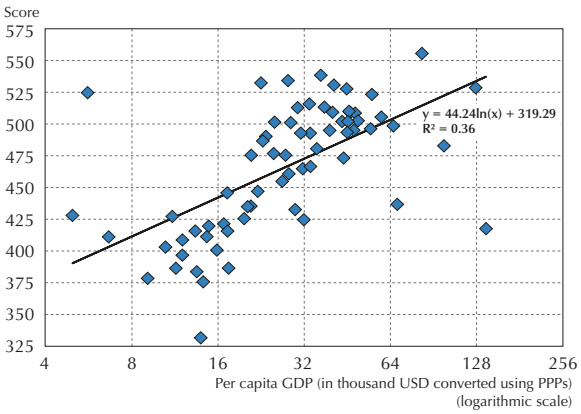
While the PISA results are representative of the target population in all adjudicated countries/economies, including Viet Nam, they cannot be readily generalised to the entire population of 15-year-olds in countries where many young people that age are not enrolled in lower or upper secondary school. Chapter 6 discusses at length the variation in coverage rates across countries and across PISA cycles. This chapter, as well as Chapters 4 and 5 about reading and mathematics performance, presents different ways to account for the share of 15-year-olds who are not covered by the PISA sample when comparing results across countries and across time.

Variations in population coverage are not the only differences that must be borne in mind when comparing results across countries. As discussed in Chapter 6, a family's wealth influences its children's performance in school, but that influence varies markedly across countries. Similarly, the relative prosperity of some countries allows them to spend more on education, while other countries find themselves constrained by a lower national income. It is therefore important to keep the national income of countries in mind when comparing the performance of education systems across countries.

Figure I.2.7 displays the relationship between national income as measured by per capita GDP and students' average science performance.[6] The figure also shows a trend line[7] that summarises the relationship between per capita GDP and mean student performance in science. The relationship suggests that 36% of the variation in countries/economies' mean scores is related to per capita GDP (23% of the variation in OECD countries). Countries with higher national incomes are thus at a relative advantage, even if the chart provides no indications about the causal nature of this relationship. This should be taken into account particularly when interpreting the performance of countries with comparatively low national income, such as Moldova and Viet Nam (Mexico and Turkey among OECD countries). Table I.2.11 shows an "adjusted" score that would be expected if the country had all of its present characteristics except that per capita GDP was equal to the average across OECD countries.

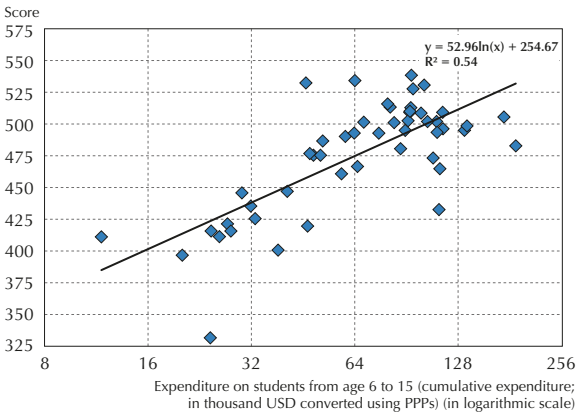
While per capita GDP reflects the potential resources available for education in each country, it does not directly measure the financial resources actually invested in education. Figure I.2.8 compares countries' actual spending per student, on average, from the age of six up to the age of 15, with average student performance in science.[8] The results are expressed in USD using purchasing power parities (PPP).

Figure I.2.7 ■ **Science performance and per capita GDP**



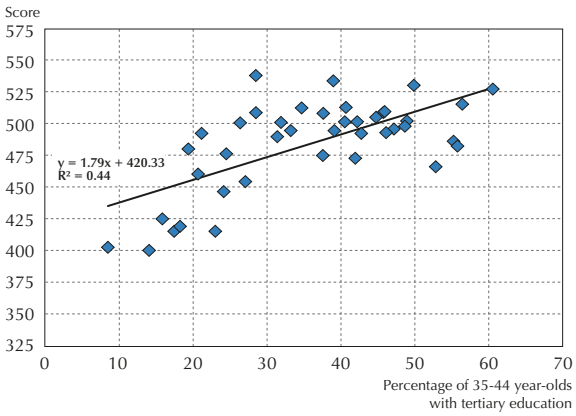
Source: OECD, PISA 2015 Database, Table I.2.11.
StatLink http://dx.doi.org/10.1787/888933431997

Figure I.2.8 ■ **Science performance and spending on education**



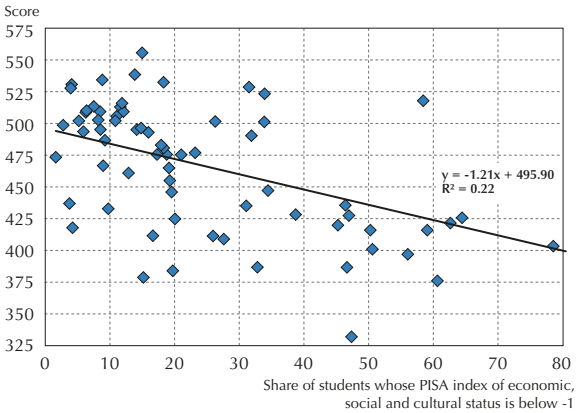
Source: OECD, PISA 2015 Database, Table I.2.11.
StatLink http://dx.doi.org/10.1787/888933432004

Figure I.2.9 ■ **Science performance and parents' education**



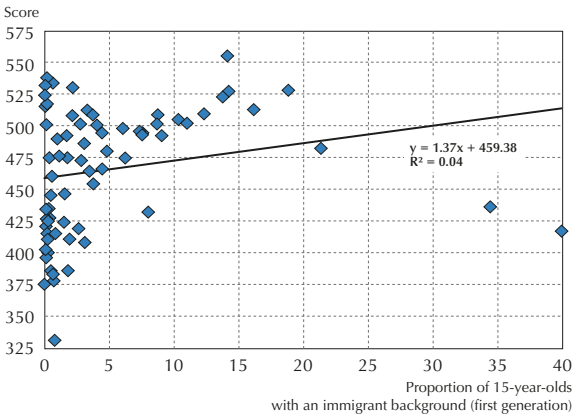
Source: OECD, PISA 2015 Database, Table I.2.11.
StatLink http://dx.doi.org/10.1787/888933432016

Figure I.2.10 ■ **Science performance and share of disadvantaged students**



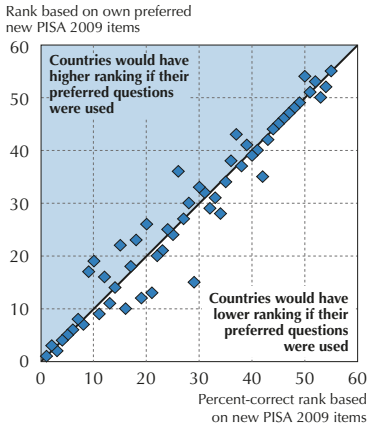
Source: OECD, PISA 2015 Database, Table I.2.11.
StatLink http://dx.doi.org/10.1787/888933432020

Figure I.2.11 ■ **Science performance and proportion of students with an immigrant background**



Source: OECD, PISA 2015 Database, Table I.2.11.
StatLink http://dx.doi.org/10.1787/888933432033

Figure I.2.12 ■ **Equivalence of the PISA assessment across cultures and languages**



Source: OECD, PISA 2012 Database, Table I.2.28, PISA 2012 Results, Volume I, http://dx.doi.org/10.1787/9789264208780-en.
StatLink http://dx.doi.org/10.1787/888933432042

Figure I.2.8 shows a positive relationship between spending per student and mean science performance. As expenditure on educational institutions per student increases, so does a country's mean performance; but the rate of increase diminishes fast, as indicated by the logarithmic scale on the horizontal axis. Expenditure per student accounts for 54% of the variation in mean performance between countries/economies (38% of the variation in OECD countries). Relatively low spending per student needs to be taken into account when interpreting the performance of countries such as Georgia and Peru (Mexico and Turkey among OECD countries). (For more details, see Figure II.6.2 in Volume II).

At the same time, deviations from the trend line suggest that moderate spending per student cannot automatically be equated with poor performance. For example, Estonia, which spends about USD 66 000 per student, and Chinese Taipei, which spends around USD 46 000 per student, perform above Austria, Luxembourg, Norway and Switzerland – all of which spend more than double this amount (more than USD 132 000 per student) (Table I.2.11).

Given the close inter-relationship between a student's performance and his or her parents' level of education, it is also important to bear in mind the educational attainment of adult populations when comparing the performance of OECD countries. Countries with more highly educated adults are at an advantage over countries where parents have less education. Figure I.2.9 shows the percentage of 35-44 year-olds who have attained tertiary education. This group corresponds roughly to the age group of parents of the 15-year-olds assessed in PISA. Parents' level of education accounts for 44% of the variation in mean performance between countries/economies (29% of the variation among OECD countries).

Socio-economic heterogeneity in student populations poses another major challenge for teachers and education systems. As shown in Chapter 6, teachers instructing socio-economically disadvantaged children are likely to face greater challenges than teachers teaching students from more advantaged backgrounds. Similarly, countries with larger proportions of disadvantaged children face greater challenges than countries with smaller proportions of these students.

Figure I.2.10 shows the proportion of students at the lower end of an international scale of the economic, social and cultural status of students, which is described in detail in Chapter 6, and how this relates to science performance. The relationship accounts for 22% of the performance variation among countries (47% of the variation among OECD countries). Among OECD countries, 64% of students in Turkey and 59% of students in Mexico belong to the most disadvantaged group, as do 34% of students in Chile and Portugal. These countries face much greater challenges than, for example, Iceland and Norway, where less than 3% of students are similarly disadvantaged (Table I.2.11). These challenges are even greater in some partner countries: 80% of students in Viet Nam and 78% of students in Indonesia are socio-economically disadvantaged.

Integrating students with an immigrant background also poses challenges to education systems (see Chapter 7). The performance of students who immigrated to the country in which they were assessed can be only partially attributed to their host country's education system. Figure I.2.11 shows the proportion of 15-year-olds with an immigrant background (excluding second-generation immigrants, who were born and educated in the country in which they were assessed) and how this relates to student performance. The relationship is positive, meaning that countries with large shares of first-generation immigrant students tend to perform better than average; but it is weak, indicating that differences in the percentage of immigrant students can, at best, account for only a small fraction of the variation in mean performance across countries.

When examining the results for individual countries, as shown in Table I.2.11, it is apparent that countries vary in their demographic, social and economic contexts. These differences need to be considered when interpreting PISA results. At the same time, the future economic and social prospects of both individuals and countries depend on the results they actually achieve, not on the performance they might have achieved under different social and economic conditions. That is why the results that are actually achieved by students, schools and countries are the focus of this volume.

Even after accounting for the demographic, economic and social context of education systems, the question remains: to what extent is an international test meaningful when differences in languages and cultures lead to very different ways in which subjects such as language, mathematics and science are taught and learned?

It is inevitable that not all tasks on the PISA assessments are equally appropriate in different cultural contexts and equally relevant in different curricular and instructional contexts. To gauge this, in 2009, PISA asked every country to identify, among the new tasks developed for use in PISA 2009, which tasks it considered most appropriate for an international test.

Countries were advised to give an on-balance rating for each task with regard to its usefulness in indicating "preparedness for life", its authenticity, and its relevance for 15-year-olds. Tasks given a high rating by a country are referred to as that country's most preferred questions for PISA. PISA then scored every country's performance on its own most preferred questions and compared the results with its performance on the entire set of new PISA tasks (see Figure I.2.12). It is clear that, in general, the proportion of questions that students answered correctly does not depend significantly on whether countries were scored only on their preferred questions or on the overall set of PISA tasks. This provides robust evidence that the results of the PISA assessments would not change markedly if countries had more influence in selecting texts that they thought might be "fairer" to their students.

STUDENTS' PROFICIENCY IN SCIENCE

PISA outcomes are reported in a variety of ways. The easiest way to summarise student performance and compare countries' relative standing in science performance is through the mean performance of students in each country. After presenting an overview of mean performance in science, this section discusses in detail the range of students' proficiency in different PISA-participating countries and economies. This range is presented in terms of the proficiency levels defined above and illustrated with sample items.

The percentage of students in each country/economy who reach each level of proficiency indicates how well countries are able to tackle underperformance while also nurturing excellence. Attaining at least Level 2 is particularly important, as Level 2 is considered a baseline level of proficiency that all young adults should be expected to attain in order to take advantage of further learning opportunities and participate fully in the social, economic and civic life of modern societies in a globalised world (OECD, 2016a; OECD, Hanushek and Woessmann, 2015).

In science, the difference between proficiency below Level 2 and proficiency at or above Level 2 corresponds to a qualitative distinction between being able to apply some limited knowledge of science in familiar contexts only (i.e. "common" knowledge), and demonstrating at least a minimum level of autonomous reasoning and understanding of the basic features of science, which, in turn, enables students to engage with science-related issues as critical and informed citizens. Students who perform below Level 2 often confuse key features of a scientific investigation, apply incorrect scientific information, and mix personal beliefs with scientific facts in support of a decision. Students who perform at or above Level 2, in contrast, can identify key features of a scientific investigation, recall single scientific concepts and information relating to a situation, and use the results of a scientific experiment represented in a data table in support of a personal decision (OECD, 2007). Education systems should strive to equip every 15-year-old with at least this basic level of proficiency in science. The percentage of students – and, more broadly, of 15-year-olds – who score at or above Level 2 on the science test indicates countries' success in achieving this goal.

Average performance in science

In 2006, the mean performance of the current 35 OECD countries was 498 score points (Table I.2.4a). In PISA 2015, the mean science score for OECD countries decreased to 493 points (an insignificant change, given the link error between the PISA 2006 and PISA 2015 scales; see the section on trends below and Annex A5). This establishes the benchmark against which each country's science performance in PISA 2015 is compared. Box I.2.1 shows how PISA score-point differences can be interpreted in terms of students' typical progression from one grade to the next.

Box I.2.1 **Interpreting differences in PISA scores: How large a gap?**

The PISA scores are represented on a scale whose units do not have a substantive meaning (unlike physical units, such as meters or grams) but are set in relation to the variation in results observed across all test participants. There is theoretically no minimum or maximum score in PISA; rather, the results are scaled to have approximately normal distributions, with means around 500 and standard deviations around 100. In statistical jargon, a one-point difference on the PISA scale therefore corresponds to an effect size of 1%; and a 10-point difference to an effect size of 10%.

A more natural, if indirect, way of representing differences in score on the PISA test is to translate scores into a grade equivalent: How far do 15-year-old students progress from one grade level to the next, in terms of PISA points?

...

Fifteen-year-old students who sit the PISA test may be enrolled in one of two or more grade levels. Based on this variation, past reports have estimated the average score-point difference across adjacent grades for countries in which a sizeable number of 15-year-olds are enrolled in at least two different grades. These estimates take into account some socio-economic and demographic differences that are also observed across grades (see Table A1.2 in OECD, 2013; 2010; 2007). On average across countries, the difference between adjacent grades is about 40 score points.

But comparisons of performance among students of the same age across different grades can only imperfectly describe how much students gain, in PISA points, over a school year. Indeed, the students who are enrolled below the expected grade for 15-year-olds differ in many ways from the students who are the same age but are enrolled in the modal grade for 15-year olds, as are those enrolled above the expected grade. Even analyses that account for differences in socio-economic and cultural status, gender and immigrant background can only imperfectly account for differences in motivation, aspirations, engagement, and many other intangible factors that influence what students know, the grade they are in, and how well they do on the PISA test.

Two types of studies can provide a better measure of the grade-equivalence of PISA scores: longitudinal follow-up studies, where the same students who took the PISA test are re-assessed later in their education, and cross-sectional designs that compare representative samples of students across adjacent age groups and grades.

In Germany, a longitudinal follow-up of the PISA 2003 cohort assessed the same 9th-grade students who participated in PISA one year later, when they were in grade 10. The comparisons showed that over this one-year period (which corresponds both to a different age and a different grade) students gained about 25 score points in the PISA mathematics test, on average, and progressed by a similar amount (21 points) in a test of science (Prenzel et al., 2006).

In Canada, the Youth in Transition Study (YITS) followed the first PISA cohort, which sat the PISA 2000 test in reading, over their further study and work career. The most recent data were collected in 2009, when these young adults were 24, and included a re-assessment of their reading score. The mean score in reading among 24-year-olds in 2009 was 598, compared to a mean score of 541 for the same young adults when they were 15 years old and in school (OECD, 2012). This shows that students continue to progress in the competencies assessed in PISA beyond age 15. At the same time, it must be borne in mind that the PISA test does not measure the more specialised kinds of knowledge and skills that young adults also acquire between the ages of 15 and 24.

In France, in 2012, 14-year-old students in grade 8 were assessed as part of a national extension to the PISA sample, at the same time as 15-year-old students who were part of the international PISA sample. The comparison of 14-year-old students in grade 8 (the modal grade for 14-year-old students in France) with students who were enrolled in the general academic track in grade 9 (15-year-old students) shows a score-point difference in mathematics of 44 points (Keskpaik and Salles, 2013). This represents an upper bound on the average progression between grades 8 and 9 in France, because some of the 14-year-olds who were included in the comparison went on to repeat grade 8 or moved to a vocational track in grade 9, and these were likely to be among the lower-performing students in that group.

Based on the PISA-based evidence cited in this box, as well as on the more general finding that learning gains on most national and international tests during one year are equal to between one-quarter and one-third of a standard deviation (Woessmann, 2016), this report equates 30 score points with about one year of schooling. This must be understood as an approximate equivalent and does not take into account national variations or differences across subjects.

When comparing mean performance across countries or across time, only those differences that are statistically significant should be taken into account (Box I.2.2 describes the different sources of uncertainty for country means and, more generally, for statistics based on PISA test results). Figure I.2.13 shows each country's/economy's mean score, and indicates for which pairs of countries/economies the differences between the means are statistically significant. For each country/economy shown in the middle column, the countries/economies whose mean scores are not statistically significantly different are listed in the right column. In all other cases, country/economy A scores higher than country/economy B if country/economy A is situated above country/economy B in the middle column, and scores lower if country/economy A is situated below country/economy B. For example: Singapore ranks first on the PISA science scale, but Japan, which appears second on the list, cannot be distinguished with confidence from Estonia and Chinese Taipei, which appear third and fourth, respectively.

In Figure I.2.13, countries and economies are divided into three broad groups: those whose mean scores are statistically around the OECD mean (highlighted in dark blue), those whose mean scores are above the OECD mean (highlighted in pale blue), and those whose mean scores are below the OECD mean (highlighted in medium blue).

Box I.2.2 When is a difference statistically significant? Three sources of statistical uncertainty

A difference is called statistically significant if it is unlikely that such a difference could be observed in the estimates based on samples, when in fact no true difference exists in the populations from which the samples are drawn.

The results of the PISA assessments for countries and economies are estimates because they are obtained from samples of students, rather than from a census of all students, and because they are obtained using a limited set of assessment tasks, not the universe of all possible assessment tasks. When students are sampled and assessment tasks are selected with scientific rigour, it is possible to determine the magnitude of the uncertainty associated with the estimate. This uncertainty needs to be taken into account when making comparisons so that differences that could reasonably arise simply due to the sampling of students and items are not interpreted as differences that actually hold for the populations. The design of the PISA test and sample are determined with respect to the objective of reducing, as much as possible, the statistical error associated with country-level statistics. Two sources of uncertainty are taken into account:

- **Sampling error:** The aim of a system-level assessment such as PISA is to generalise the results based on samples to the larger target population. The sampling methods used in PISA ensure not only that the samples are representative and provide a valid estimate of the population mean score and distribution, but also that the error due to sampling is reduced to a minimum. The sampling error decreases with the number of schools and (to a lesser extent) of students included in the assessment. The sampling error associated with a country's mean performance estimate is, for most countries, around 2 to 3 PISA score points. For the OECD average (which is based on 35 independent national samples) the sampling error is reduced to about 0.4 PISA score point.
- **Measurement error** (also called imputation error): No test is perfect and can fully measure broad concepts such as science literacy. The use of a limited number of items to assess broad domains, for instance, introduces some measurement uncertainty: would the use of a different set of items have resulted in different performance? This uncertainty is quantified in PISA. Among other things, it decreases with the number of items in a domain that underlie a proficiency estimate. It is therefore somewhat larger for minor domains than for major domains, and it is larger for individual students (who only see a fraction of all test items) than for country means (which are based on all test items). It also decreases with the amount of background information available. For country mean estimates, the imputation error is smaller than the sampling error (around 0.5 PISA score point).

When comparing results across different PISA cycles an additional source of uncertainty must be taken into account. Indeed, even if different PISA assessments use the same metric for measuring performance (for science, this metric was defined in PISA 2006, when science was, for the first time, the major focus of the PISA test), the test instruments and items used in the assessment change in each cycle, as do the calibration samples and sometimes the statistical models used for scaling results. To make the results directly comparable over time, scales have to be equated; this means that results are transformed so that they can be expressed on the same metric. The *link error* quantifies the uncertainty around the equating of scales. The procedures used for equating PISA 2015 results to prior scales are described in Annex A5; further details on the link error and the equating procedures are provided in the *PISA 2015 Technical Report* (OECD, forthcoming).

The link error affects all scaled values equally and is therefore independent of the size of the student sample. As a result, it is the same for estimates based on individual countries, on subpopulations, or on the OECD average. For comparisons between science results in PISA 2015 and science results in PISA 2006, the link error corresponds to about 4.5 score points, making it by far the most significant source of uncertainty in trend comparisons.

Figure I.2.13 ■ Comparing countries' and economies' performance in science

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

Mean score	Comparison country/ economy	Countries and economies whose mean score is NOT statistically significantly different from the comparison country's/economy's score
556	Singapore	
538	Japan	Estonia, Chinese Taipei
534	Estonia	Japan, Chinese Taipei, Finland
532	Chinese Taipei	Japan, Estonia, Finland, Macao (China), Canada, Viet Nam
531	Finland	Estonia, Chinese Taipei, Macao (China), Canada, Viet Nam
529	Macao (China)	Chinese Taipei, Finland, Canada, Viet Nam, Hong Kong (China)
528	Canada	Chinese Taipei, Finland, Macao (China), Viet Nam, Hong Kong (China), B-S-J-G (China)
525	Viet Nam	Chinese Taipei, Finland, Macao (China), Canada, Hong Kong (China), B-S-J-G (China), Korea
523	Hong Kong (China)	Macao (China), Canada, Viet Nam, B-S-J-G (China), Korea
518	B-S-J-G (China)	Canada, Viet Nam, Hong Kong (China), Korea, New Zealand, Slovenia, Australia, United Kingdom, Germany, Netherlands
516	Korea	Viet Nam, Hong Kong (China), B-S-J-G (China), New Zealand, Slovenia, Australia, United Kingdom, Germany, Netherlands
513	New Zealand	B-S-J-G (China), Korea, Slovenia, Australia, United Kingdom, Germany, Netherlands
513	Slovenia	B-S-J-G (China), Korea, New Zealand, Australia, United Kingdom, Germany, Netherlands
510	Australia	B-S-J-G (China), Korea, New Zealand, Slovenia, United Kingdom, Germany, Netherlands, Switzerland
509	United Kingdom	B-S-J-G (China), Korea, New Zealand, Slovenia, Australia, Germany, Netherlands, Switzerland, Ireland
509	Germany	B-S-J-G (China), Korea, New Zealand, Slovenia, Australia, United Kingdom, Netherlands, Switzerland, Ireland
509	Netherlands	B-S-J-G (China), Korea, New Zealand, Slovenia, Australia, United Kingdom, Germany, Switzerland, Ireland
506	Switzerland	Australia, United Kingdom, Germany, Netherlands, Ireland, Belgium, Denmark, Poland, Portugal, Norway
503	Ireland	United Kingdom, Germany, Netherlands, Switzerland, Belgium, Denmark, Poland, Portugal, Norway, United States
502	Belgium	Switzerland, Ireland, Denmark, Poland, Portugal, Norway, United States
502	Denmark	Switzerland, Ireland, Belgium, Poland, Portugal, Norway, United States
501	Poland	Switzerland, Ireland, Belgium, Denmark, Portugal, Norway, United States, Austria, Sweden
501	Portugal	Switzerland, Ireland, Belgium, Denmark, Poland, Norway, United States, Austria, France, Sweden
498	Norway	Switzerland, Ireland, Belgium, Denmark, Poland, Portugal, United States, Austria, France, Sweden, Czech Republic, Spain
496	United States	Ireland, Belgium, Denmark, Poland, Portugal, Norway, Austria, France, Sweden, Czech Republic, Spain, Latvia
495	Austria	Poland, Portugal, Norway, United States, France, Sweden, Czech Republic, Spain, Latvia
495	France	Portugal, Norway, United States, Austria, Sweden, Czech Republic, Spain, Latvia
493	Sweden	Poland, Portugal, Norway, United States, Austria, France, Czech Republic, Spain, Latvia, Russia
493	Czech Republic	Norway, United States, Austria, France, Sweden, Spain, Latvia, Russia
493	Spain	Norway, United States, Austria, France, Sweden, Czech Republic, Latvia, Russia
490	Latvia	United States, Austria, France, Sweden, Czech Republic, Spain, Russia
487	Russia	Sweden, Czech Republic, Spain, Latvia, Luxembourg, Italy, CABA (Argentina)
483	Luxembourg	Russia, Italy, CABA (Argentina)
481	Italy	Russia, Luxembourg, Hungary, Lithuania, Croatia, CABA (Argentina)
477	Hungary	Italy, Lithuania, Croatia, CABA (Argentina), Iceland
475	Lithuania	Italy, Hungary, Croatia, CABA (Argentina), Iceland
475	Croatia	Italy, Hungary, Lithuania, CABA (Argentina), Iceland
475	CABA (Argentina)	Russia, Luxembourg, Italy, Hungary, Lithuania, Croatia, Iceland, Israel, Malta
473	Iceland	Hungary, Lithuania, Croatia, CABA (Argentina), Israel
467	Israel	CABA (Argentina), Iceland, Malta, Slovak Republic
465	Malta	CABA (Argentina), Israel, Slovak Republic
461	Slovak Republic	Israel, Malta, Greece
455	Greece	Slovak Republic, Chile, Bulgaria
447	Chile	Greece, Bulgaria
446	Bulgaria	Greece, Chile, United Arab Emirates
437	United Arab Emirates	Bulgaria, Uruguay, Romania, Cyprus[1]
435	Uruguay	United Arab Emirates, Romania, Cyprus[1]
435	Romania	United Arab Emirates, Uruguay, Cyprus[1], Moldova, Albania, Turkey
433	Cyprus[1]	United Arab Emirates, Uruguay, Romania, Moldova, Albania, Turkey
428	Moldova	Romania, Cyprus[1], Albania, Turkey, Trinidad and Tobago, Thailand
427	Albania	Romania, Cyprus[1], Moldova, Turkey, Trinidad and Tobago, Thailand
425	Turkey	Romania, Cyprus[1], Moldova, Albania, Trinidad and Tobago, Thailand, Costa Rica, Qatar
425	Trinidad and Tobago	Moldova, Albania, Turkey, Thailand
421	Thailand	Moldova, Albania, Turkey, Trinidad and Tobago, Costa Rica, Qatar, Colombia, Mexico
420	Costa Rica	Turkey, Thailand, Qatar, Colombia, Mexico
418	Qatar	Turkey, Thailand, Costa Rica, Colombia, Mexico
416	Colombia	Thailand, Costa Rica, Qatar, Mexico, Montenegro, Georgia
416	Mexico	Thailand, Costa Rica, Qatar, Colombia, Montenegro, Georgia
411	Montenegro	Colombia, Mexico, Georgia, Jordan
411	Georgia	Colombia, Mexico, Montenegro, Jordan
409	Jordan	Montenegro, Georgia, Indonesia
403	Indonesia	Jordan, Brazil, Peru
401	Brazil	Indonesia, Peru
397	Peru	Indonesia, Brazil
386	Lebanon	Tunisia, FYROM
386	Tunisia	Lebanon, FYROM
384	FYROM	Lebanon, Tunisia
378	Kosovo	Algeria
376	Algeria	Kosovo
332	Dominican Republic	

1. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus."

Source: OECD, PISA 2015 Database, Table I.2.3.

StatLink http://dx.doi.org/10.1787/888933432052

Twenty-four countries and economies perform above the OECD average in science. One country, Singapore, outperforms all other countries and economies in science, with a mean score of 556 points. Japan (538 points) scores below Singapore, but above all other countries, except Estonia (534 points) and Chinese Taipei (532 points), whose mean scores are not statistically significantly different. Together with Japan and Estonia, Finland (531 points) and Canada (528 points) are the four highest-performing OECD countries. The mean scores in Macao (China) (529 points), Viet Nam (525 points), Hong Kong (China) (523 points) and Beijing-Shanghai-Jiangsu-Guangdong (China) (hereafter "B-S-J-G [China]") (518 points), as well as in OECD countries Korea (516 points), New Zealand and Slovenia (513 points each), Australia (510 points), Germany, the Netherlands and the United Kingdom (509 points each), Switzerland (506 points), Ireland (503 points), Belgium and Denmark (502 points each), Poland and Portugal (501 points each), and Norway (498 points) also lie above the OECD average.

Countries that perform around the average include Austria, the Czech Republic, France, Latvia, Spain, Sweden and the United States. Thirty-nine participating countries and economies score below the OECD average.

The gap in performance between the highest- and the lowest-performing OECD countries is 123 score points. That is, while the average score of the highest-performing OECD country, Japan (538), is about half a standard deviation above the OECD average (the equivalent of more than one year of schooling; see Box I.2.1), the average score of the lowest-performing OECD country, Mexico (416 points), is more than three-quarters of a standard deviation, or the equivalent of more than two years of schooling, below the OECD average. But the performance difference observed among partner countries and economies is even larger, with a 224 score-point difference between Singapore (556 points) and the Dominican Republic (332 points).

Because the figures are derived from samples, and because of the statistical uncertainty associated with mean estimates, it is not possible to determine a country's/economy's precise ranking among all participating countries and economies. However, it is possible to identify, with 95% confidence, a range of rankings in which the country's/economy's performance level lies (Figure I.2.14). This range of ranks can be wide, particularly for countries/economies whose scores are similar to those of many other countries/economies. For example, the United States ranks between 21st and 31st among all countries/economies (between 15th and 25th among OECD countries only).

For subnational entities whose results are reported in Annex B2, a rank order was not estimated; but the mean score and its confidence interval allow for a comparison of performance with that of countries and economies. For example, Alberta (Canada) and British Columbia (Canada) show a score just below that of top-performer Singapore and similar to that of Japan.

Students at the different levels of proficiency in science

Figure I.2.15 shows the distribution of students at each of the seven proficiency levels. The percentage of students performing below Level 2 is shown on the left side of the vertical axis.

Proficiency above the baseline

Proficiency at Level 2 (scores higher than 410 but lower than 484 points)

At Level 2, students can draw on everyday content knowledge and basic procedural knowledge to identify an appropriate scientific explanation, interpret data, and identify the question being addressed in a simple experimental design. They can use common scientific knowledge to identify a valid conclusion from a simple data set. Level 2 students demonstrate basic epistemic knowledge by being able to identify questions that could be investigated scientifically.

Question 2 from the unit *METEOROIDS AND CRATERS* (Annex C1) is typical of Level 2 tasks. It asks a simple question about the relationship between a planet's atmosphere and the likelihood that meteoroids will burn up before hitting the planet's surface. The question focuses on the ability to make a correct prediction ("The thicker a planet's atmosphere is, the fewer craters its surface will have because more meteoroids will burn up in the atmosphere"), based on knowledge of earth and space systems. It is therefore categorised as a question requiring the competence of explaining phenomena scientifically, based on content knowledge, related to earth and space systems.

To answer the question correctly, students must demonstrate some basic knowledge about earth and space systems. The short introductory text provides numerous cues to help students identify the correct relationship ("Rocks in space that enter Earth's atmosphere are called meteoroids. Meteoroids heat up, and glow as they fall through Earth's atmosphere. Most meteoroids burn up before they hit Earth's surface."). Question 3B in the same unit is another Level 2 task related to the same categories. In contrast to Question 2, students are not given any cue, but the knowledge required to solve this question is familiar and simple.

Figure I.2.14 [Part 1/2] ■ Science performance among PISA 2015 participants, at national and subnational levels

	Science scale					
			Range of ranks			
			OECD countries		All countries/economies	
	Mean score	95% confidence interval	Upper rank	Lower rank	Upper rank	Lower rank
Singapore	556	553 - 558			1	1
Alberta (Canada)	541	533 - 549				
British Columbia (Canada)	539	530 - 547				
Japan	538	533 - 544	1	2	2	3
Quebec (Canada)[1]	537	528 - 546				
Estonia	534	530 - 538	1	3	2	5
Chinese Taipei	532	527 - 538			2	7
Finland	531	526 - 535	2	4	3	7
Massachusetts (United States)	529	516 - 542				
Macao (China)	529	526 - 531			5	8
Canada	528	524 - 532	3	4	5	9
Viet Nam	525	517 - 532			4	10
Ontario (Canada)	524	516 - 532				
Hong Kong (China)	523	518 - 528			7	10
Castile and Leon (Spain)	519	512 - 526				
B-S-J-G (China)	518	509 - 527			8	16
Nova Scotia (Canada)	517	508 - 526				
Korea	516	510 - 522	5	8	9	14
Madrid (Spain)	516	509 - 523				
Flemish community (Belgium)	515	510 - 521				
Bolzano (Italy)	515	511 - 520				
Prince Edward Island (Canada)	515	504 - 525				
New Zealand	513	509 - 518	5	9	10	15
Slovenia	513	510 - 515	5	9	11	15
England (United Kingdom)	512	506 - 518				
Navarre (Spain)	512	504 - 520				
Galicia (Spain)	512	506 - 518				
Trento (Italy)	511	506 - 515				
Australia	510	507 - 513	6	11	12	17
United Kingdom	509	504 - 514	6	13	12	19
Germany	509	504 - 514	6	13	12	19
Netherlands	509	504 - 513	7	13	13	19
Aragon (Spain)	508	498 - 517				
New Brunswick (Canada)	506	498 - 515				
Newfoundland and Labrador (Canada)	506	500 - 512				
Switzerland	506	500 - 511	8	17	14	23
German-speaking community (Belgium)	505	496 - 515				
Catalonia (Spain)	504	495 - 513				
Ireland	503	498 - 507	11	18	17	24
Lombardia (Italy)	503	493 - 512				
North Carolina (United States)	502	493 - 512				
Belgium	502	498 - 506	12	19	18	25
Denmark	502	497 - 507	12	19	18	25
Poland	501	497 - 506	12	19	18	25
Asturias (Spain)	501	494 - 509				
Portugal	501	496 - 506	12	19	18	25
Northern Ireland (United Kingdom)	500	495 - 506				
Manitoba (Canada)	499	490 - 509				
Norway	498	494 - 503	14	21	20	27
La Rioja (Spain)	498	487 - 509				
Castile-La Mancha (Spain)	497	490 - 505				
Scotland (United Kingdom)	497	492 - 501				
United States	496	490 - 502	15	25	21	31
Saskatchewan (Canada)	496	490 - 502				
Cantabria (Spain)	496	485 - 507				
Austria	495	490 - 500	17	24	23	30
France	495	491 - 499	18	24	24	30
Comunidad Valenciana (Spain)	494	488 - 500				
Sweden	493	486 - 500	18	25	24	32

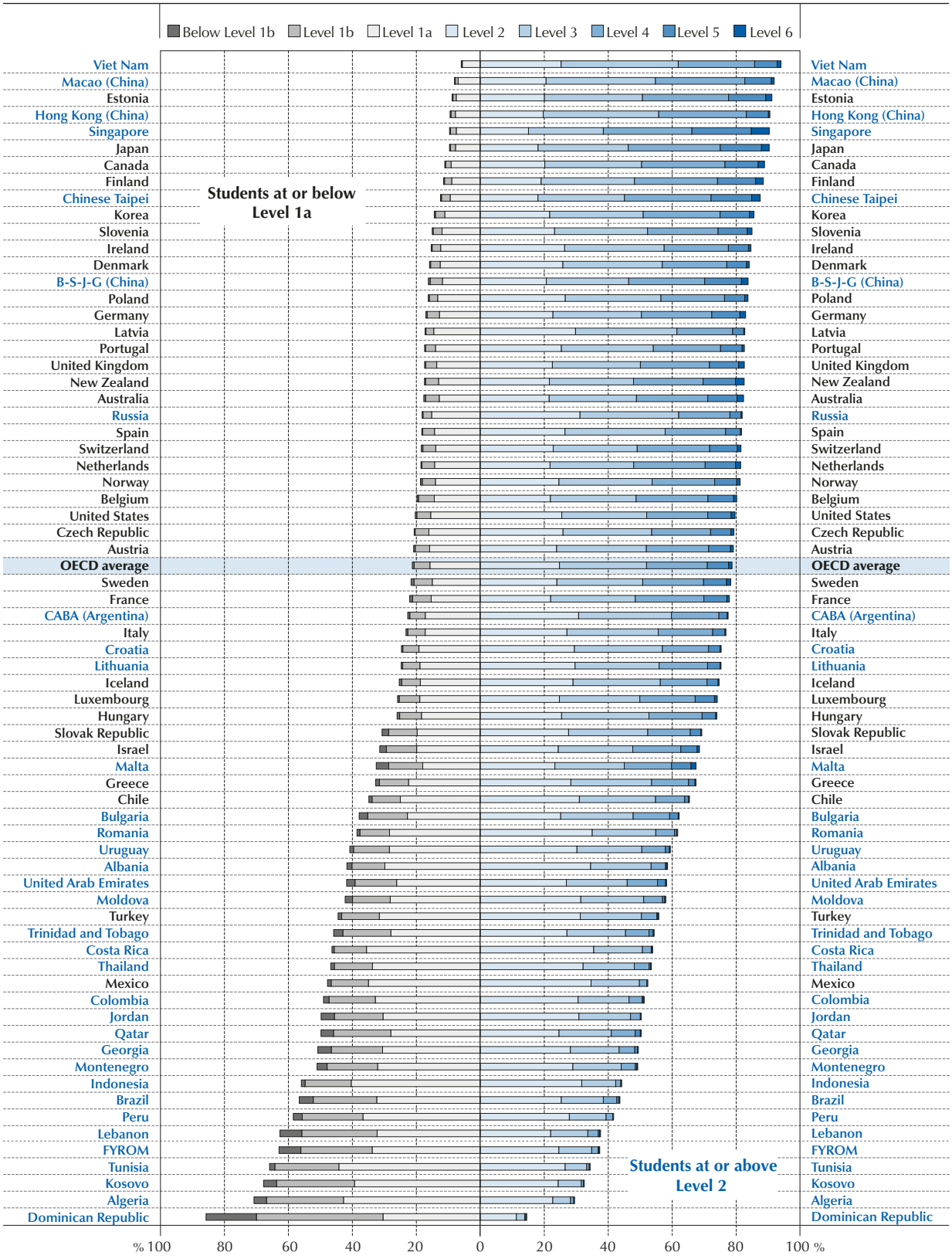
* See note 1 under Figure I.2.13.
1. Results for the province of Quebec in this table should be treated with caution due to a possible non-response bias.
2. Puerto Rico is an unincorporated territory of the United States. As such, PISA results for the United States do not include Puerto Rico.
Note: OECD countries are shown in bold black. Partner countries, economies and subnational entities that are not included in national results are shown in bold blue.
Regions are shown in black italics (OECD countries) or blue italics (partner countries).
Countries and economies are ranked in descending order of mean science performance.
Source: OECD, PISA 2015 Database.
StatLink http://dx.doi.org/10.1787/888933432060

Figure I.2.14 [Part 2/2] ■ Science performance among PISA 2015 participants, at national and subnational levels

	Science scale					
			Range of ranks			
			OECD countries		All countries/economies	
	Mean score	95% confidence interval	Upper rank	Lower rank	Upper rank	Lower rank
Czech Republic	493	488 - 497	19	25	25	31
Spain	493	489 - 497	20	25	25	31
Latvia	490	487 - 493	23	25	28	32
Russia	487	481 - 492			30	34
French community (Belgium)	485	477 - 494				
Balearic Islands (Spain)	485	476 - 493				
Wales (United Kingdom)	485	479 - 490				
Murcia (Spain)	484	476 - 491				
Basque Country (Spain)	483	477 - 489				
Luxembourg	483	481 - 485	26	27	32	34
Italy	481	476 - 485	26	28	32	36
Dubai (UAE)	480	477 - 483				
Hungary	477	472 - 481	27	29	34	39
Lithuania	475	470 - 481			34	39
Canary Islands (Spain)	475	468 - 482				
Croatia	475	471 - 480			35	39
CABA (Argentina)	475	463 - 487			32	41
Extremadura (Spain)	474	467 - 482				
Iceland	473	470 - 477	28	29	36	39
Andalusia (Spain)	473	465 - 481				
Região Autónoma dos Açores (Portugal)	470	465 - 474				
Israel	467	460 - 473	30	31	39	42
Malta	465	462 - 468			40	42
Slovak Republic	461	456 - 466	30	32	41	43
Bogotá (Colombia)	458	448 - 467				
Greece	455	447 - 463	31	32	42	44
Chile	447	442 - 452	33	33	44	45
Bulgaria	446	437 - 454			43	46
Campania (Italy)	445	435 - 455				
United Arab Emirates	437	432 - 441			46	49
Uruguay	435	431 - 440			46	49
Romania	435	429 - 441			46	50
Manizales (Colombia)	434	426 - 443				
Medellín (Colombia)	433	425 - 442				
Cyprus*	433	430 - 435			47	50
Sharjah (UAE)	432	414 - 451				
Moldova	428	424 - 432			49	53
Albania	427	421 - 434			49	54
Turkey	425	418 - 433	34	34	49	55
Trinidad and Tobago	425	422 - 427			51	54
Abu Dhabi (UAE)	423	414 - 432				
Thailand	421	416 - 427			51	57
Cali (Colombia)	421	412 - 430				
Costa Rica	420	416 - 424			53	57
Qatar	418	416 - 420			55	58
Colombia	416	411 - 420			55	60
Mexico	416	412 - 420	35	35	55	59
Montenegro	411	409 - 413			59	61
Georgia	411	406 - 416			58	61
Jordan	409	403 - 414			59	62
Indonesia	403	398 - 408			61	63
Puerto Rico[2]	403	391 - 415				
Ajman (UAE)	402	395 - 408				
Fujairah (UAE)	401	391 - 412				
Brazil	401	396 - 405			62	64
Ras Al Khaimah (UAE)	400	384 - 417				
Peru	397	392 - 401			63	64
Umm Al Quwain (UAE)	387	379 - 395				
Lebanon	386	380 - 393			65	67
Tunisia	386	382 - 391			65	67
FYROM	384	381 - 386			65	67
Kosovo	378	375 - 382			68	69
Algeria	376	371 - 381			68	69
Dominican Republic	332	327 - 337			70	70

* See note 1 under Figure I.2.13.
1. Results for the province of Quebec in this table should be treated with caution due to a possible non-response bias.
2. Puerto Rico is an unincorporated territory of the United States. As such, PISA results for the United States do not include Puerto Rico.
Note: OECD countries are shown in bold black. Partner countries, economies and subnational entities that are not included in national results are shown in bold blue.
Regions are shown in black italics (OECD countries) or blue italics (partner countries).
Countries and economies are ranked in descending order of mean science performance.
Source: OECD, PISA 2015 Database.
StatLink http://dx.doi.org/10.1787/888933432060

Figure I.2.15 ■ **Students' proficiency in science**



Countries and economies are ranked in descending order of the percentage of students who perform at or above Level 2.
Source: OECD, PISA 2015 Database, Table I.2.1a.

StatLink http://dx.doi.org/10.1787/888933432072

Level 2 is considered the baseline level of science proficiency that is required to engage in science-related issues as a critical and informed citizen. Indeed, the baseline level of proficiency defines the level of achievement on the PISA scale at which students begin to demonstrate the science competencies that will enable them to participate effectively and productively in life situations related to science and technology. More than 90% of students in Viet Nam (94.1%), Macao (China) (91.9%), Estonia (91.2%), Hong Kong (China) (90.6%), and Singapore and Japan (both 90.4%) meet this benchmark. Across OECD countries, an average of 79% of students attains Level 2 or higher; more than one in two students in all OECD countries perform at these levels (Figure I.2.15 and Table I.2.1a).

In many middle- and low-income countries, many 15-year-olds are not eligible to participate in PISA because these young people have dropped out of school, never attended school, or are in school, but in grade 6 or below (see Chapter 6). Assuming that these 15-year-olds would not reach Level 2 if they sat the PISA science test, and based on the estimated total number of 15-year-olds in each country/economy, it is possible to estimate the proportion of all 15-year-olds who reach a baseline level of performance in science.

Similar assumptions of below-baseline skills among the population of 15-year-olds not covered by PISA are often made in related literature (UNESCO, 2004; Hanushek and Woessmann, 2008; Spaull and Taylor, 2015; Taylor and Spaull, 2015).[9] The PISA pilot initiative to survey out-of-school children in five countries, which will be implemented in 2017 (see Box I.6.3 in Chapter 6), will provide first-of-its-kind data on the reading and mathematics skills of this population in relation to the international PISA scale. In the absence of similar data for all PISA-participating countries, the hypothesis of below-baseline skills provides a lower bound on the percentage of 15-year-olds who are proficient above the baseline level.

In 22 countries and economies, including OECD countries Mexico and Turkey, as well as Viet Nam, whose mean performance in PISA is above the OECD average, fewer than one in two 15-year-olds is in school, in grade 7 or above, and reaches at least Level 2 on the PISA science scale. In Viet Nam, 94% of students who are in the PISA target population attain Level 2; but the PISA target population represents less than 50% of the overall population of 15-year-olds. In Algeria, the Dominican Republic, Kosovo and Lebanon, fewer than one in four 15-year-olds reaches this level of proficiency in science (Figure I.2.16 and Table I.2.1b).

Proficiency at Level 3 (scores higher than 484 but lower than 559 points)

At Level 3, students can draw upon moderately complex content knowledge to identify or construct explanations of familiar phenomena. In less familiar or more complex situations, they can construct explanations with relevant cueing or support. They can draw on elements of procedural or epistemic knowledge to carry out a simple experiment in a constrained context. Level 3 students are able to distinguish between scientific and non-scientific issues and identify the evidence supporting a scientific claim.

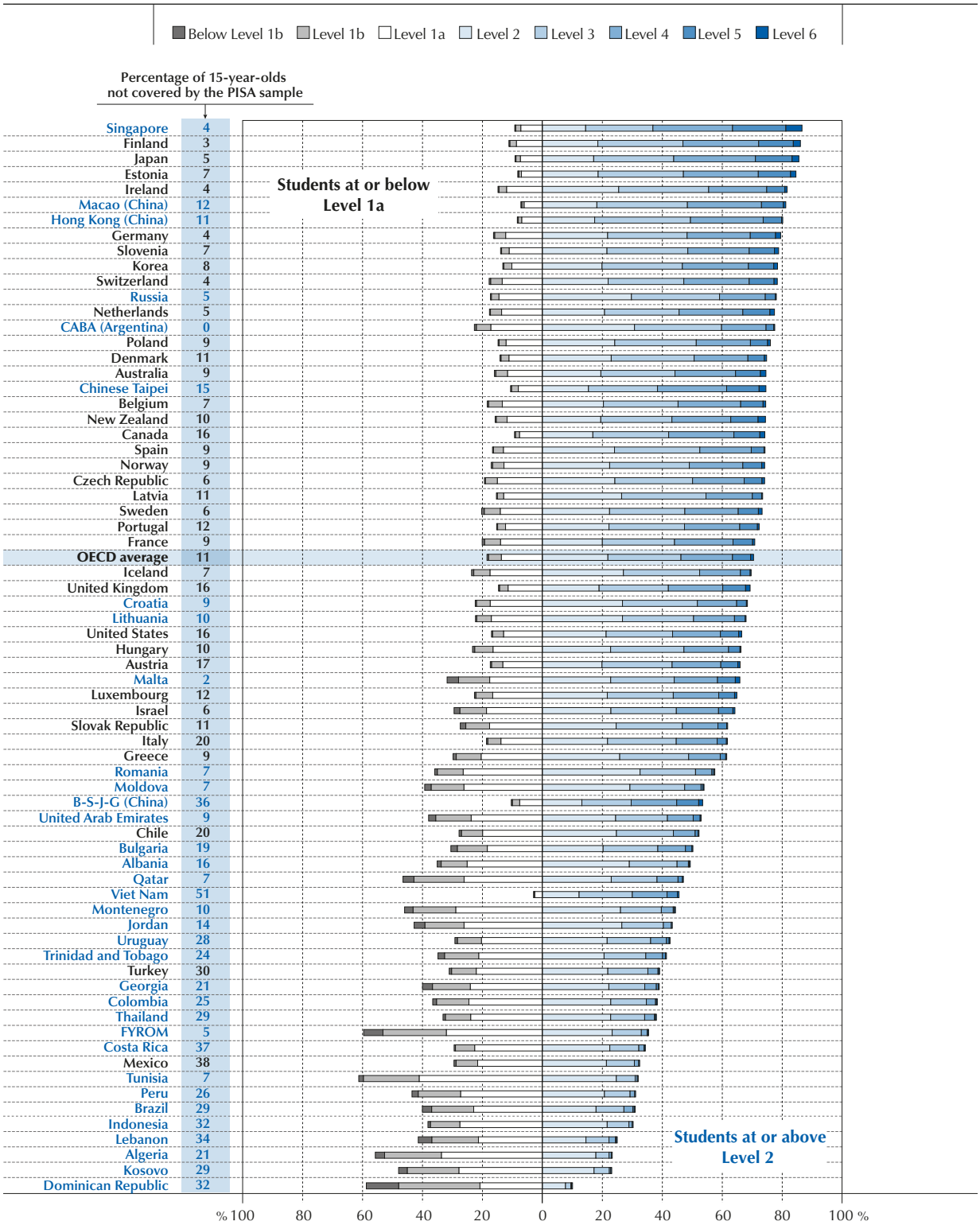
An example of a question at Level 3 is Question 1 in *BIRD MIGRATION* (Annex C1). Similar to the two questions used to illustrate proficiency at Level 2, this question requires the competency to explain phenomena scientifically based on content knowledge – in this case, a basic knowledge of the theory of evolution. The question states that most bird species migrate in large groups, rather than individually, and that this behaviour is the result of evolution. In order to answer this question correctly, students must identify which of the four possible explanations is consistent with the theory of evolution and with the observed facts: that birds that migrated individually or in small groups were less likely to survive and have offspring.

Question 1 in unit *SLOPE-FACE INVESTIGATION* is also a Level 3 task. In the introduction, test-takers are presented with the observation that there is a dramatic difference in the vegetation on the two slopes of a valley. The first question then presents the design used by a group of students for collecting data about the conditions that prevail on the two slopes. Students are asked to evaluate this design (the question is classified as "evaluating and designing scientific enquiry"), and to explain the rationale behind it. This is an open-ended question, where test-takers' answers must demonstrate epistemic knowledge – in this case, knowledge of (at least one) rationale for taking multiple, independent measurements in order to identify how conditions vary across the two slopes.

In most OECD countries, Level 3 corresponds to a median level of performance. The median score, i.e. the score that divides the population in two equal halves – one scoring above the median, and one below – falls within Level 3. On average across OECD countries, more than half of all students (54.0%) are proficient at Level 3 or higher (that is, at Level 3, 4, 5 or 6). Similarly, Level 3 corresponds to the median proficiency of students in 31 participating countries and economies. Across OECD countries on average, 27.2% of students score at Level 3, the largest share among the seven proficiency levels described in PISA. Similarly, in 31 countries and economies, the largest share of students performs at Level 3 (Figure I.2.15 and Table I.2.1a).

Figure I.2.16 ■ **Fifteen-year olds' proficiency in science**

Students at the different levels of proficiency in science, as a percentage of all 15-year-olds



Note: The length of each bar is proportional to the percentage of 15-year-olds covered by the PISA sample (Coverage index 3; see Annex A2).

Countries and economies are ranked in descending order of the number of students who perform at or above Level 2, expressed as a percentage of the total population of 15-year-olds in the country.

Source: OECD, PISA 2015 Database, Table I.2.1b.

StatLink http://dx.doi.org/10.1787/888933432083

Proficiency at Level 4 (scores higher than 559 but lower than 633 points)

At Level 4, students can use more sophisticated content knowledge, which is either provided or recalled, to construct explanations of more complex or less familiar events and processes. They can conduct experiments involving two or more independent variables in a constrained context. They can justify an experimental design, drawing on elements of procedural and epistemic knowledge. Level 4 students can interpret data drawn from a moderately complex data set or less familiar contexts and draw appropriate conclusions that go beyond the data and provide justifications for their choices.

Question 2 in unit *SLOPE-FACE INVESTIGATION* (Annex C1), which typifies a Level 4 question, requires students to evaluate two claims by interpreting the provided data (it is classified as "interpreting data and evidence scientifically"). The data include confidence intervals around the average of measurements of solar radiation, soil moisture and rainfall. Students are asked to demonstrate an understanding of how measurement error affects the degree of confidence associated with specific scientific measurements, one major aspect of epistemic knowledge. Question 2 in unit *BIRD MIGRATION* is located at the top of Level 4 (630 points on the PISA scale). It is an example of a question where students must draw on procedural knowledge to identify a factor that could result in an inadequate or inaccurate set of data, and explain its effect on the quality of scientific enquiry. Both tasks exemplify the more complex knowledge and more sophisticated understanding demonstrated by students who are proficient at Level 4, compared to students at the lower levels of proficiency.

On average across OECD countries, 26.7% of students perform at Level 4 or above, and score higher than 559 points on the PISA science scale. The largest share of students in Japan, Singapore and Chinese Taipei performs at this level (modal level); and Level 4 is the median level of performance in Singapore, where 51.9% of students score at or above this level (Figure I.2.15 and Table I.2.1a).

Proficiency at Level 5 (scores higher than 633 but lower than 708 points)

At Level 5, students can use abstract scientific ideas or concepts to explain unfamiliar and more complex phenomena, events and processes. They can apply more sophisticated epistemic knowledge to evaluate alternative experimental designs, justify their choices and use theoretical knowledge to interpret information or make predictions. Students at this level can evaluate ways of exploring a given question scientifically and identify limitations in interpretations of data sets, including sources and the effects of uncertainty in scientific data.

There are no released items from the PISA 2015 main survey to illustrate proficiency at Level 5 (although, as noted, Question 2 in unit *BIRD MIGRATION* is located near the limit between Level 4 and Level 5). Question 5 in the field trial unit *RUNNING IN HOT WEATHER* (Annex C1), however, presents an example of tasks that students at this level are typically able to solve. It requires students to use their knowledge of biology (content knowledge) to explain the role of sweating in regulating the body's temperature. This is a complex phenomenon due to the indirect nature of the effects; the requirement to provide the answer in an open text entry field also contributes to difficulty.

Level 5 on the science scale marks another qualitative difference. Students who can complete Level 5 tasks can be said to be top performers in science in that they are sufficiently skilled in and knowledgeable about science to be able to creatively and autonomously apply their knowledge and skills to a wide variety of situations, including unfamiliar ones.

On average across OECD countries, 7.7% of students are top performers, meaning that they are proficient at Level 5 or 6. About one in four (24.2%) students in Singapore, and just under one in six students in Chinese Taipei (15.4%) and Japan (15.3%) performs at this level. In 11 countries/economies (Australia, Canada, B-S-J-G [China], Estonia, Finland, Germany, Korea, the Netherlands, New Zealand, Slovenia and the United Kingdom), between 10% and 15% of all students perform at Level 5 or above. By contrast, in 20 countries/economies, including OECD countries Turkey (0.3%) and Mexico (0.1%), fewer than one in 100 students is a top performer (Figure I.2.15 and Table I.2.1a).

Proficiency at Level 6 (scores higher than 708 points)

Students at Level 6 on the PISA science scale can successfully complete the most difficult items in the PISA science assessment. At Level 6, students can draw on a range of interrelated scientific ideas and concepts from the physical, life, and earth and space sciences and use procedural and epistemic knowledge to offer explanatory hypotheses of novel scientific phenomena, events and processes that require multiple steps, or to make predictions. In interpreting data and evidence, they can discriminate between relevant and irrelevant information and can draw on knowledge external to the normal school curriculum. They can distinguish between arguments that are based on scientific evidence and theory and those based on other considerations. Level 6 students can evaluate competing designs of complex experiments, field studies or simulations, and justify their choices.

Question 1 in the example unit *SUSTAINABLE FISH FARMING* (Annex C1) requires Level 6 proficiency. This question requires students to understand an ecosystem (here, a fish farm) and the role of several organisms within that system. The main competency required is to explain phenomena scientifically. In order to answer correctly, students must understand the goal of the fish farm, the function of each of the three tanks therein, and which organisms will best fulfill each function. Students must use information provided in the stimulus and the diagram, including a footnote under the diagram. An additional component that adds difficulty is the open-ended nature of the task. Any of the four organisms can be placed in any of the three tanks and there is no restriction on the number of organisms in each tank. As a result, there are multiple ways of getting this incorrect. The issue of sustainable fish farming is in the "living systems" content area, and the solution of this item mainly draws on content knowledge.

On average across OECD countries, 1.1% of students attain Level 6. Singapore has the largest proportion of students (5.6%) who score at this level in science. In New Zealand and Chinese Taipei, 2.7% of students score at Level 6 in science. In 18 participating countries and economies, between one in 40 (2.5%) and one in 100 (1%) students score at this level, while in 49 other countries/economies, fewer than one in 100 students scores at the highest level (Figure I.2.15 and Table I.2.1a).

Proficiency below the baseline

Proficiency at Level 1a (scores higher than 335 but lower than 410 points)

At Level 1a, students can use common content and procedural knowledge to recognise or identify explanations of simple scientific phenomenon. With support, they can undertake structured scientific enquiries with no more than two variables. They can identify simple causal or correlational relationships and interpret graphical and visual data that require a low level of cognitive ability. Students at Level 1a can select the best scientific explanation for given data in familiar personal, local and global contexts.

There are no released items from the PISA 2015 main survey to illustrate proficiency at Level 1a. Paper-based questions developed for the PISA 2006 assessment of science can be used to illustrate the competencies of students who score at this Level (OECD, 2009).

Across OECD countries, 15.7% of students perform at Level 1a, and only 5.5% of students perform below Level 1a. In the Dominican Republic, fewer than one in two students (about 45%) attains this (or a higher) level of performance. In 17 countries and economies, including OECD countries Mexico and Turkey, the largest share of students performs at this level (Figure I.2.15 and Table I.2.1a).

Proficiency at Level 1b (scores higher than 261 but lower than 335 points)

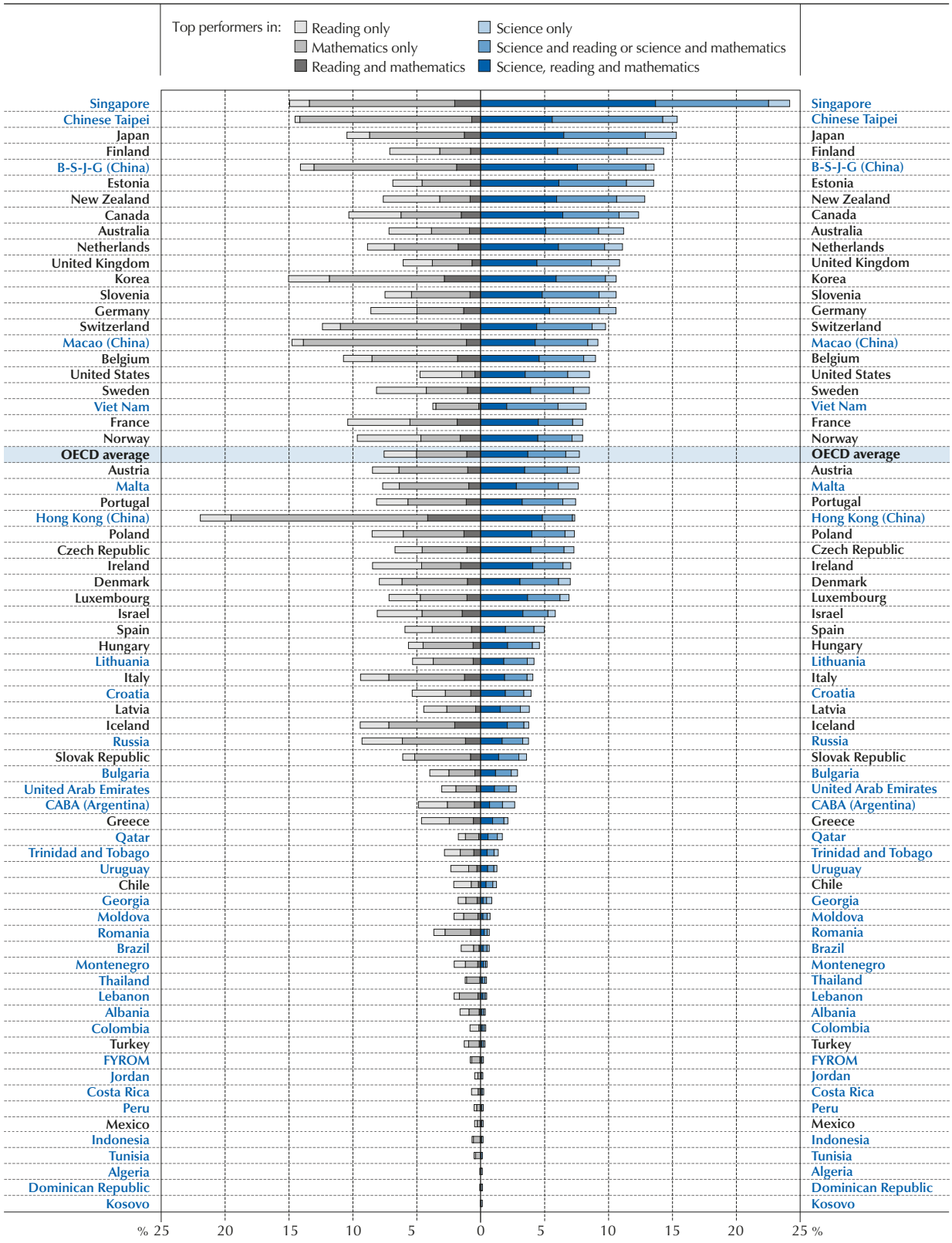
At Level 1b, students can use common content knowledge to recognise aspects of simple scientific phenomena. They can identify simple patterns in data, recognise basic scientific terms and follow explicit instructions to carry out a scientific procedure.

Question 3A in the unit *METEOROIDS AND CRATERS* (Annex C1) is an example of a task at Level 1b. In order to solve this question, students must use common scientific knowledge to match the size of a meteoroid with the size of the crater it would create on a planet's surface, based on an image showing three craters of different sizes. Since it is common knowledge that a larger object would cause a larger crater and a smaller one would cause a smaller crater, the question is located at the bottom of the "interpret data and evidence scientifically" scale.

Across OECD countries, 4.9% of students perform at Level 1b and 0.6% performs below Level 1b. In 40 countries and economies, including Canada, Estonia, Hong Kong (China), Japan, Macao (China) and Viet Nam, less than 10% of students perform at or below Level 1b; in those six countries, less than 2% of students perform at this level (Figure I.2.15 and Table I.2.1a).

No item in the PISA assessment can indicate what students who perform below Level 1b can do. Students below Level 1b may have acquired some elements of science knowledge and skills, but based on the tasks included in the PISA test, their ability can only be described in terms of what they cannot do – and they are unlikely to be able to solve, other than by guessing, any of the PISA tasks. In some countries, the proportion of students who perform below Level 1b is substantial: 15.8% in the Dominican Republic, and between 4% and 7% in Lebanon, FYROM, Brazil, Georgia, Jordan and Kosovo (in descending order of that proportion).

Figure I.2.17 ■ **Overlapping of top performers in science with top performers in reading and mathematics**



Countries and economies are ranked in descending order of the percentage of top performers in science only and in science with other domains.
Source: OECD, PISA 2015 Database, Table I.2.9a.

StatLink http://dx.doi.org/10.1787/888933432092

Where are the top performers in science?

Performance in PISA is measured by students' ability to complete increasingly complex tasks. Only a small proportion of students attains the highest levels of proficiency – Level 5 or 6 – and can be called top performers in science, reading or mathematics. Even fewer students are academic all-rounders: those who achieve proficiency Level 5 or higher in all three subjects. These students can draw on and use information from multiple and indirect sources to solve complex problems, and can integrate knowledge from across different areas. Such exceptional skills can provide a significant advantage in a competitive, knowledge-based global economy.

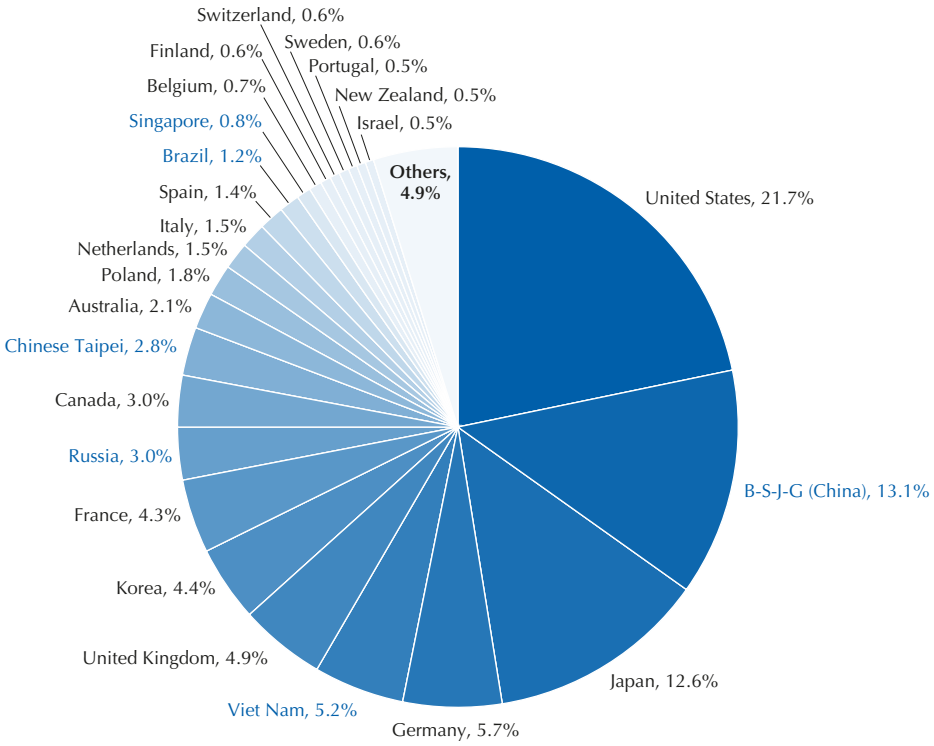
Figure I.2.17 shows the proportion of top performers in science and all-rounders across PISA-participating countries and economies. The parts of the diagram shaded in blue represent the percentage of 15-year-old students who are top performers in science, with darker tones for top-performing students in science who also excel at similar levels in reading and/or mathematics. The grey parts to the left of the diagram show the percentage of 15-year-old students who are top performers in mathematics and/or reading but not in science.

Figure I.2.18 depicts the number of 15-year-old students who are proficient at Level 5 or 6 on the PISA science scale, by country. While Figure I.217 shows the share of students in each country who perform at Level 5 or 6, it does not take into account that the student population varies in size across countries. Yet both the proportion of top performers within a country and the size of countries matter when establishing countries' contributions to the global pool of top-performing students. Even though the proportion of top performers in science is comparatively small in the United States, the United States represents a fifth of the total shown in Figure I.2.18 (which, of course, considers only countries participating in PISA), simply because of the size of the country and the overall number of 15-year-old students that the PISA sample represents.

In contrast, Singapore, which has the largest share of 15-year-olds performing at Level 5 or 6 on the PISA science scale, contributes less than 1% to the global pool of top-performing students because its population is relatively small.

Figure I.2.18 ■ **The global pool of top performers: A PISA perspective**

Proportion of all PISA top performers in science in individual countries/economies



Source: OECD, PISA 2015 Database, Table I.2.9c.

StatLink http://dx.doi.org/10.1787/888933432102

As shown in Figure I.2.18, more than half of all top-performing students in PISA live in just four countries/economies: the United States (22%), B-S-J-G (China) (13%), Japan (13%) and Germany (6%). Ten countries/economies are home to over 75% of the global pool of top performers in science, as measured by PISA. In addition to the four countries with the largest talent pool listed above, the United Kingdom and Viet Nam each contribute 5%, France and Korea about 4%, and Canada and Russia about 3% to the global pool of top-performing students. When considered together, the 35 OECD countries represent 72% of the global pool of top-performing students, and the 28 European Union members represent 26% of that pool (Table I.2.9c).

GENDER DIFFERENCES IN SCIENCE PERFORMANCE

Table I.2.7 presents a summary of boys' and girls' performance on the PISA science assessment. On average across OECD countries, boys' mean performance in science is 4 points higher than girls' – a statistically significant, but numerically small difference. Boys score significantly above girls, on average, in 24 countries and economies. The largest advantage for boys is found in Austria, Costa Rica and Italy, where the difference between boys' and girls' scores is over 15 points. Girls score significantly above boys, on average, in 22 countries and economies. In Albania, Bulgaria, Finland, FYROM, Georgia, Jordan, Qatar, Trinidad and Tobago, and the United Arab Emirates, girls' mean score is more than 15 score points higher than boys'.

In general, boys show greater variation in performance than girls. In all but 18 countries and economies (where the difference is not significant), the variation in science performance (measured by the standard deviation) is larger among boys than among girls (Table I.2.7). As a result, on average across OECD countries, the share of top-performing students (those who perform at or above Level 5) is larger among boys than among girls, but so is the share of low-achieving students (those who perform below Level 2 on the science scale). Whereas 8.9% of boys perform at or above Level 5, only 6.5% of girls perform at that level (Figure I.2.20). At the same time, 21.8% of boys do not reach a baseline level of proficiency in science, a slightly larger proportion than that of girls (20.7%) (Figure I.2.19).

In 33 countries and economies, the share of top performers in science is larger among boys than among girls (Figure I.2.20). Among the countries where more than 1% of students are top performers in science, in Austria, Chile, Ireland, Italy, Portugal and Uruguay, around two out of three top-performing students are boys. Finland is the only country in which there are significantly more girls than boys among top performers.

Boys are over-represented compared to girls among low-achieving students in science in 28 countries/economies, while girls are over-represented in 5 countries/economies (Figure I.2.19). In the remaining countries/economies, the gender difference in the share of low performers and top performers is not statistically significant.

TRENDS IN STUDENTS' SCIENCE PERFORMANCE

PISA 2015 is the sixth round of PISA since the programme was launched in 2000. Every PISA test assesses students' science, reading and mathematics literacy; in each round, one of these subjects is the main domain and the other two are minor domains (see "What is PISA?" at the beginning of this volume).

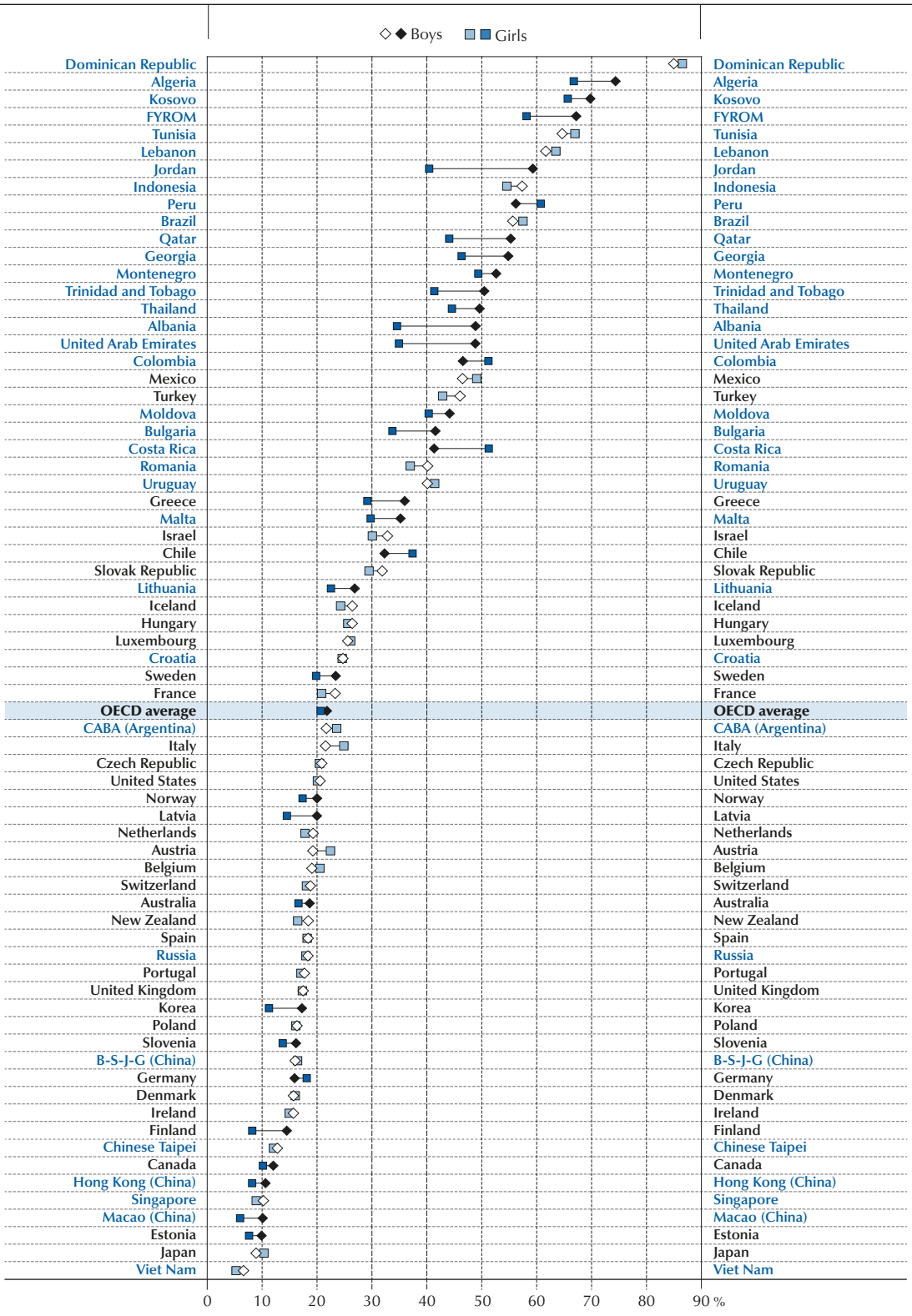
The first full assessment of each domain sets the scale and starting point for future comparisons. Science was the major domain for the first time in 2006, and is again the major domain in PISA 2015. This means that it is possible to measure the change in science performance between PISA 2015 and any prior PISA test, starting with PISA 2006, but not with respect to PISA 2000 or 2003. The most reliable way to establish a trend for science performance is to compare all available results between 2006 and 2015.

Trends in student performance indicate whether and how school systems are improving. Trends in science performance are available for 64 countries and economies that participated in PISA 2015. Fifty-one of these have science performance data for 2015 and data from the three previous comparable PISA assessments (2006, 2009 and 2012); five have data from 2015 and two additional assessments; and eight countries and economies have data from 2015 and one previous assessment.

To better understand a country's /economy's trends and maximise the number of countries in the comparisons, this report focuses on the average three-year trend in student performance. The three-year trend is the average rate of change observed over three-year intervals during the available period (three years correspond to the typical interval between two PISA assessments; the magnitude of the average three-year trend can therefore be directly compared to the change observed between two consecutive assessments, e.g. PISA 2012 and PISA 2015). For countries and economies that have participated in all four PISA assessments, the average three-year trend takes into account all four points in time; for those countries that have valid data for fewer assessments, the average three-year trend takes into account only the valid and available information.

Figure I.2.19 ■ **Gender differences among low-achieving students in science**

Percentage of boys and girls performing below Level 2 in science

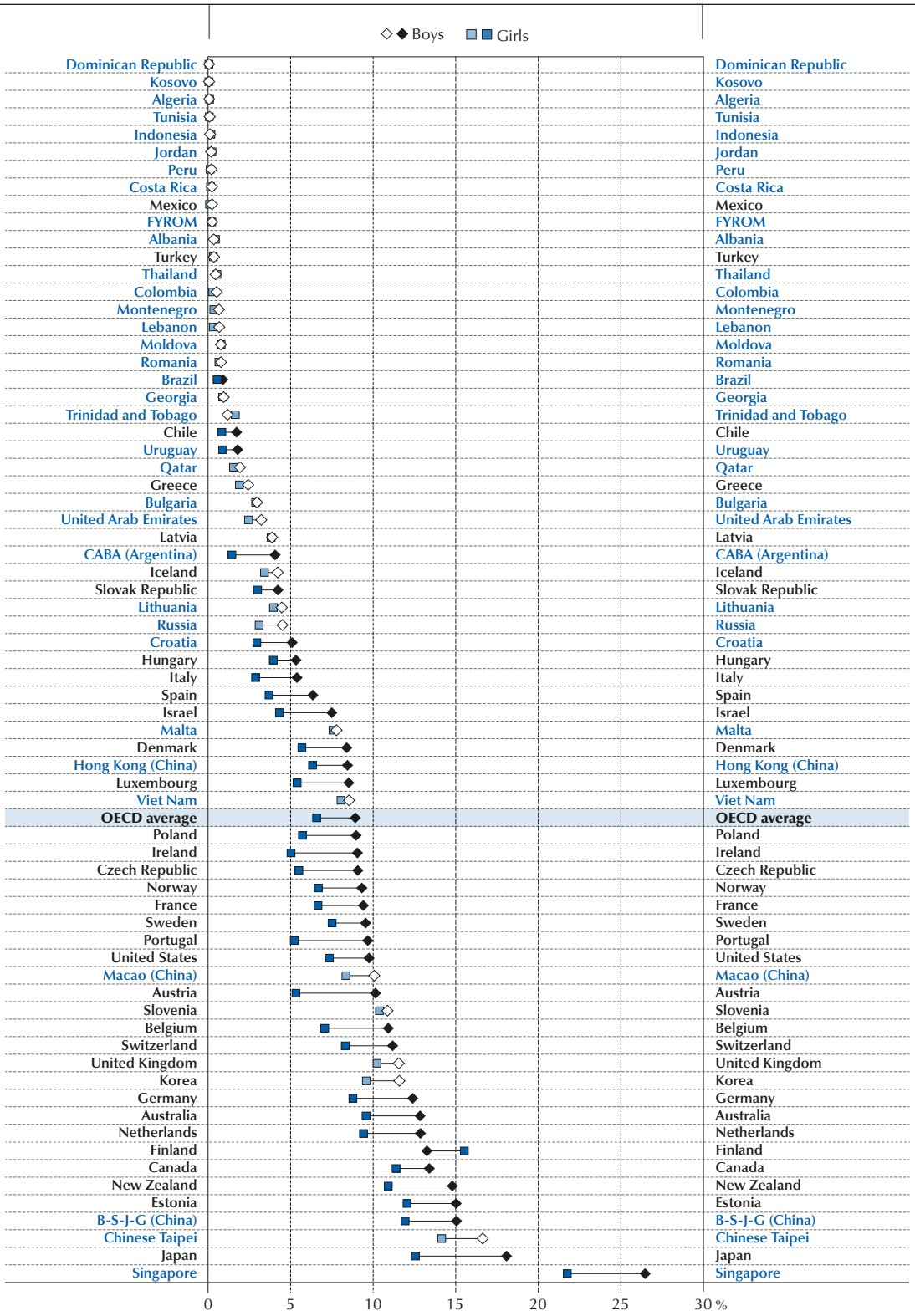


Note: Statistically significant differences between boys and girls are marked in a darker tone (see Annex A3).
Countries and economies are ranked in descending order of the percentage of low-achieving boys.
Source: OECD, PISA 2015 Database, Table I.2.6a.

StatLink http://dx.doi.org/10.1787/888933432113

Figure I.2.20 ■ **Gender differences among top performers in science**

Percentage of boys and girls performing at or above Level 5 in science



Note: Statistically significant differences between boys and girls are marked in a darker tone (see Annex A3).
Countries and economies are ranked in ascending order of the percentage of top-performing boys.
Source: OECD, PISA 2015 Database, Table I.2.6a.

StatLink http://dx.doi.org/10.1787/888933432129

The methodologies underpinning the analysis of performance trends in international studies of education are complex (see Annex A5). In order to ensure the comparability of successive PISA results, a number of conditions must be met.

First, successive assessments must include a sufficient number of common assessment items so that results can be reported on a common scale. The set of items included must adequately cover the different aspects of the framework for each domain. Because the results of Kazakhstan in 2015 are based only on multiple-choice items, they cannot be reliably compared to the results of other countries, nor to Kazakhstan's results in previous assessments (see Annex A4 for details).

Second, the sample of students in successive assessments must be equally representative of the target population, and only results from samples that meet the strict standards set by PISA can be compared over time. Even though they participated in successive PISA assessments, some countries and economies cannot compare all their PISA results over time. For example, the PISA 2015 sample for Malaysia did not meet the PISA response-rate standards, so comparisons with 2015 cannot be reported for Malaysia. The PISA 2015 sample for Argentina did not cover the full target population, due to the potential omission of schools from the sampling frame, except for the adjudicated region of Ciudad Autónoma de Buenos Aires (Argentina) (hereafter "CABA [Argentina]"); as a result, only results for CABA (Argentina) can be compared over time (see Annex A4 for details).

Even when PISA samples accurately reflect the target population (that of 15-year-olds enrolled in grade 7 and above), changes in enrolment rates and demographics can affect the interpretation of trends. To distinguish between changes that affect equivalent populations and changes related to the composition of the target population, adjusted trends that account for population changes are presented in addition to the basic measure of performance change across PISA samples.

Third, the assessment conditions must be sufficiently similar across time so that performance on the test reflects the same underlying proficiency in a domain.[10] Ensuring the equivalence of trend items across time is particularly important in the context of PISA 2015, when most countries/economies that participated in the assessment conducted the test on computer (see Box I.2.3 and Annex A5).

Box I.2.3 Can past PISA results in science be compared to results from the computer-based PISA 2015 science test?

PISA aims to measure, at each point in time, the knowledge and skills that are required to participate fully in society and the economy. Because these evolve slowly over time, every nine years PISA revisits the framework and the instruments used to measure the domains of reading, mathematics and science. This periodic revision of frameworks and instruments also provides an opportunity to align PISA with new developments in assessment techniques and with the latest understanding of the cognitive processes underlying proficiency in each domain.

The PISA 2015 assessment coincided with the development of an updated framework for science, the major domain, and with the development of new items to capture all aspects of this updated framework. The existing items (trend items) that were used in PISA 2006, 2009 and 2012 were also reviewed against this updated framework.

A major difference with previous assessments of science is the delivery of test questions on computers. Most of the countries/economies participating in the PISA 2015 test, including all OECD countries, assessed their students on computers (see "What is PISA" at the beginning of this volume). In order to compare the results of this test to those obtained by earlier cohorts of students on past PISA paper-based tests, it was necessary to establish first the equivalence of the paper- and the computer-based instruments (Janssen, 2011).

Paper and computer tests in PISA are linked through common items (so-called "link items", or "link tasks"); all of these items were developed, initially, for the paper-based tests in previous PISA rounds. The PISA 2015 field trial tested the equivalence of link items between computer-based tests and paper-based tests. Two levels of equivalence were distinguished: scalar (strong) and metric (weak) equivalence (Davidov, Schmidt and Billet, 2011; Meredith, 1993). Only items that passed the test of equivalence were retained for the main study; among these, a majority of items (61 out of 85 in science) attained the highest level of invarianceand were used as link items for science.

Comparing current PISA scores to past PISA scores, or PISA scores in one country to PISA scores in another country, is supported by a large number of link items that attain the highest level of equivalence (scalar invariance). Annex A5 and the *PISA 2015 Technical Report* (OECD, forthcoming) provide details about the number of scalar invariant items for other domains and about the mode-effect study conducted in the context of the PISA 2015 field trial.

Fourth, the same reporting scale must be used to report student proficiency. In PISA, the reporting scale is re-estimated in each cycle, and then equated to the scale constructed the first time a domain became the major domain. The uncertainty associated with equating scales is included when computing the significance of changes or trend estimates (see Box I.2.2). PISA 2015 introduced several changes in the scaling of the test. Annex A5 describes the technical details of these changes, and how they affect trend comparisons.

In addition, not all countries have participated in all PISA assessments. When computing the OECD average changes and trends in science performance, only those countries with valid data to compare among assessments are included in the average. While comparisons between the 2006 and 2015 results in science use data from all 35 OECD member countries, only 34 OECD countries can compare their 2009 and 2015 results. For this reason, tables and figures showing trends in science performance often include two distinct averages – the OECD average-35, which includes all OECD countries, and the OECD average-34, which excludes Austria.

Average three-year trend in performance

The average three-year trend is used as the main measure of trends in a country's/economy's science, reading and mathematics performance. The average three-year trend for the mean is the average rate at which a country's/economy's mean score in mathematics, reading and science has changed over consecutive three-year periods throughout its participation in PISA assessments. Similarly, the average three-year trend for the median (the score that divides a population in two equal halves – one scoring above the median, and one below) is the average rate at which a country's/economy's median score in mathematics, reading and science has changed over consecutive three-year periods throughout its participation in PISA assessments. The interval of three years is chosen to correspond to the usual interval between two PISA assessments. Thus, a positive average three-year trend of x points indicates that the country/economy has improved in performance by x points on average in each PISA assessment since its earliest comparable PISA results. For countries and economies that have participated in only two assessments, the average three-year trend is equal to the score-point difference between the two assessments, divided by the number of years that passed between the assessments and multiplied by three.

The average three-year trend is a more robust measure of a country's/economy's progress in education outcomes than the simple difference between two points in time as it is based on information available from all assessments. For countries that participated in more than two PISA assessments, it is thus less sensitive to statistical fluctuations that may alter a country's/economy's trends in PISA performance if results are compared between only two assessments. This robustness comes at the cost of ignoring accelerations, decelerations or reversals of the rate of change: the average three-year trend assumes that the rate of change is steady over the period considered (linear trend). The average three-year trend also takes into account the fact that, for some countries and economies, the period between PISA assessments is less than three years. This is the case for those countries and economies that participated in PISA 2009 as part of PISA+: they conducted the assessment in 2010 instead of 2009.

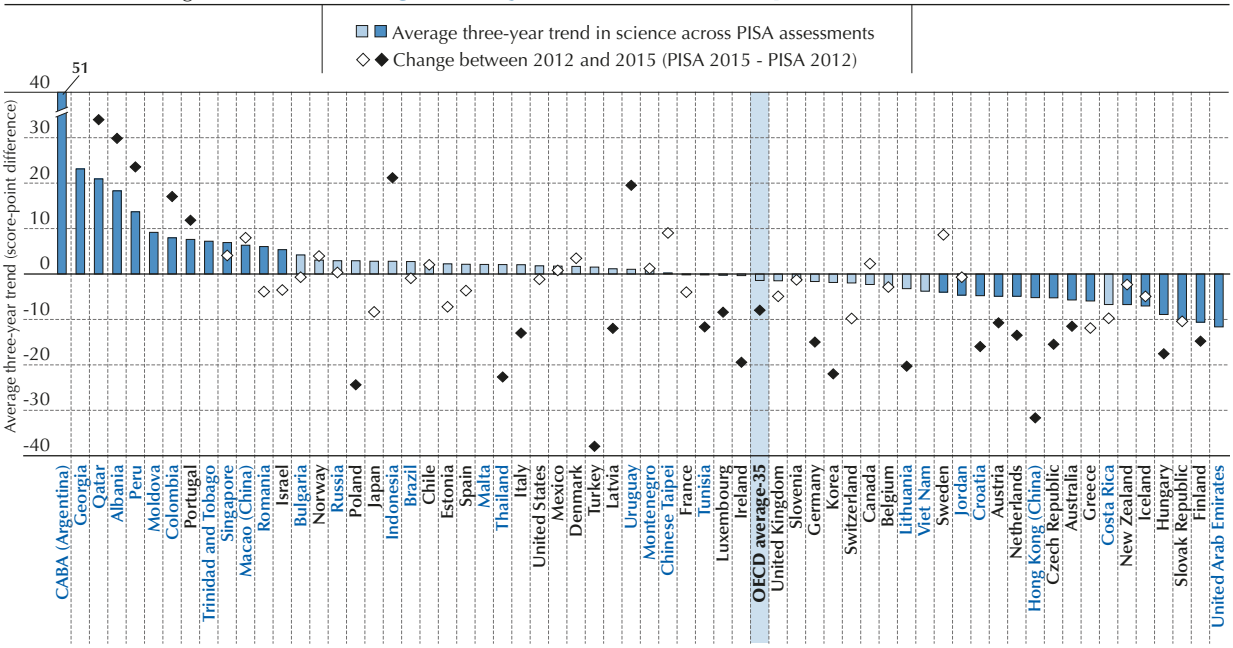
Table I.2.4a shows the average three-year trend in mean science performance. Table I.2.4b presents the three-year trend for the 10th, 25th, 75th and 90th percentiles, as well as for the median (50th percentile) in science performance.

On average across OECD countries with comparable data in PISA 2006 and PISA 2015, performance has remained stable (a non-significant decline of 1.4 points every 3 years was observed). But the stability of the average masks the significant changes observed in many countries and economies. Of the 64 countries/economies with valid results in more than one PISA round, about half (31) show no significant change in mean performance, 15 countries show a significant average improvement in science performance, and 18 show a significant average deterioration in performance.

As Figure I.2.21 shows, in CABA (Argentina), Georgia and Qatar, student performance in science improved by more than 20 score points every 3 years since these countries/economies began participating in PISA (however, Georgia only participated in PISA 2009 and PISA 2015, and CABA [Argentina] participated as a separate adjudicated entity since only PISA 2012). Albania, Moldova and Peru improved by between 9 and 20 score points every 3 years since 2009, and Colombia improved by 8 points, on average, every 3 years throughout its participation in PISA (since 2006).

Among OECD countries, improvements in mean science performance are observed in Portugal (with an average improvement of more than seven score points every three years), Israel (about five score points every three years), Norway and Poland (about three score points every three years). Partner countries/economies Macao (China), Romania, Singapore, and Trinidad and Tobago also show significant improvements over the period in which they participated in PISA. (Of these countries and economies, only Macao [China] and Romania participated in all four PISA cycles between 2006 and 2015.)

Figure I.2.21 ■ **Average three-year trend in science performance since 2006**



Notes: Statistically significant differences are shown in a darker tone (see Annex A3).

The average three-year trend is the average rate of change, per three-year period, between the earliest available measurement in PISA and PISA 2015. For countries and economies with more than one available measurement, the average three-year trend is calculated with a linear regression model. This model takes into account that Costa Rica, Georgia, Malta and Moldova conducted the PISA 2009 assessment in 2010 as part of PISA 2009+. For countries/economies with comparable data for PISA 2012 and PISA 2015 only, the average three-year trend coincides with the change between 2012 and 2015.

Only countries/economies with valid results for PISA 2015 and at least one prior assessment are shown.

Countries and economies are ranked in descending order of the average three-year trend in science performance.

Source: OECD, PISA 2015 Database, Table I.2.4a.

StatLink http://dx.doi.org/10.1787/888933432133

Among the 15 countries and economies that have a negative average three-year trend, 13 have comparable data for all four assessments between PISA 2006 and PISA 2015, the United Arab Emirates did not participate until PISA 2012, and results for PISA 2009 in Austria cannot be compared with previous or later assessments (see note 9 at the end of this chapter). In Finland, the Slovak Republic and the United Arab Emirates, student performance in science deteriorated by more than 10 points every three years, on average (i.e. assuming a steady rate of change). Performance in Australia, the Czech Republic, Greece, Hong Kong (China), Hungary, Iceland and New Zealand deteriorated between five and ten points every three years; and mean performance in science in Austria, Croatia, Jordan, the Netherlands and Sweden declined by less than five points every three years on average.

Change in science performance between 2012 and 2015

For countries that participated in both PISA 2012 and PISA 2015, Figure I.2.21 also displays the change in PISA results over the most recent period. By contrasting the change over the three years from 2012 to 2015, indicated by the diamonds, and the average three-year trend over a longer period of time, indicated by the bars, it is possible to assess whether a country's/economy's improvement or deterioration over the most recent period confirms, or contradicts, the trend observed over a longer period of time. For countries that have valid data only in PISA 2012 and PISA 2015, the two values coincide, and diamonds are therefore not shown; but in general, when more than two assessments are available, the two do not necessarily coincide, and long-term trends are more precisely estimated than short-term changes. On average across OECD countries, performance was similar in 2015 and 2006, but significantly lower (by eight score points) in 2015 than in 2012.

Among countries/economies with a significant, negative trend, in Croatia, the Czech Republic and Hong Kong (China), average science scores decreased over the most recent period more than 10 points faster than the average rate of change over PISA assessments, indicating an acceleration or inversion of the trend observed between 2006 and 2012. By contrast, in Sweden, the most recent period shows a non-significant improvement of nine points. This reflects a deceleration, or perhaps inversion, of the negative trend observed over the longer period.

Among countries with a significant, positive trend, in Albania and Qatar, mean science scores improved between 2012 and 2015 more than 10 points faster than the average rate of change over PISA cycles, indicating a possible acceleration of the trend.

Some countries/economies that show no significantly positive or negative trend, on average, nevertheless show a significant improvement, or deterioration, over the most recent period. Germany, Ireland, Italy, Korea, Latvia, Lithuania, Luxembourg, Poland, Thailand, Tunisia and Turkey,[11] for example, all have significantly lower mean scores in 2015 than in 2012. Meanwhile, Indonesia and Uruguay have a significantly higher score in 2015 than in 2012, but show no significant average improvement over a longer period of time.

Average three-year trend in performance, accounting for changes in enrolment rates

Changes in a country's or economy's science performance can have many sources. In some countries, a decline in mean performance may result from a lower quality of education than in the past. But in other cases, a similar decline may, in fact, reflect an improvement in the capacity of education systems to include students who would not have attended school in previous years, or who, at age 15, would still have been in primary school. Changes can also result from demographic shifts in the country's population. By following strict sampling and methodological standards, PISA ensures that all countries and economies measure the science performance of their 15-year-old students in grades 7 and above; but because of changes in enrolment rates, migration or other demographic and social trends, the characteristics of this reference population may change.

Adjusted trends neutralise some of the changes observed in the composition and coverage of the PISA sample so that it becomes possible to identify some of the sources of the trends observed. In this volume, two types of adjusted trends are presented. The first accounts for changes in enrolment rates over time, and is presented in this section. The second accounts for changes in the age (measured in quarters), gender, and immigrant background, and is presented in the next section. Annex A5 provides details on how these adjusted trends were calculated.

Over the past 10 years, many countries – particularly low- and middle-income countries – have made great efforts to ensure that every child completes primary school (at least), and to reduce dropout rates in secondary education. Some countries, such as Brazil and Turkey, have raised the age at which students can leave compulsory education to over 15; and these reforms have been accompanied by a significant increase in the share of 15-year-olds who are included in the PISA target population. This expansion in education opportunities makes it more difficult to interpret the observed trends in performance for the countries concerned.

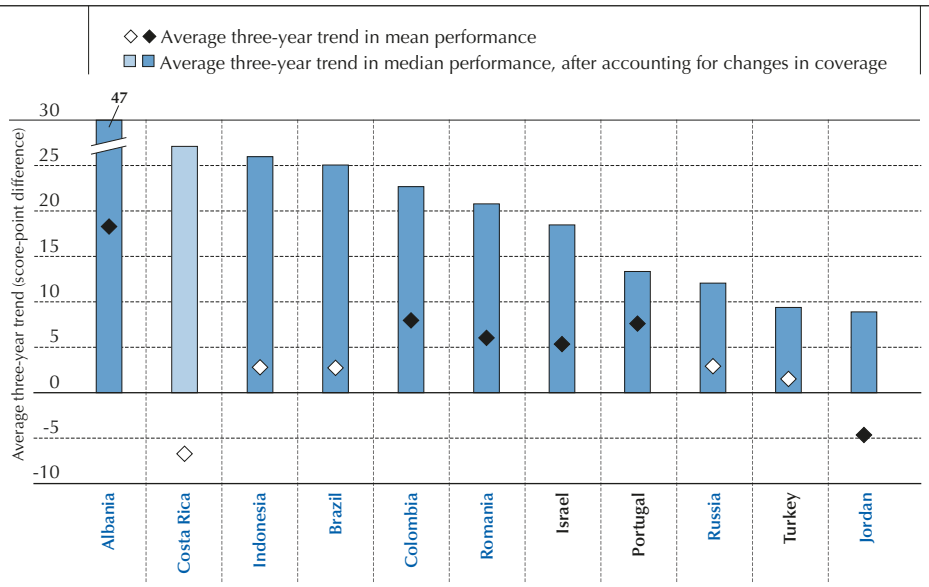
It is impossible to know for certain what the PISA score of the 15-year-olds who were not enrolled in school or who were still in grades 1 through 6 would have been, had they been tested. Without attributing an exact score to these students, it is nevertheless possible to assume, with some confidence, that they would have scored in the bottom half of a country's performance distribution (see Hanushek and Woessmann, 2008; Spaull and Taylor, 2015; Taylor and Spaull, 2015; as well as note 8 at the end of this chapter for related assumptions). Given this assumption, it is possible to track, over time, the change in the median performance of 15-year-olds in a country – i.e. the minimum level achieved by at least 50% of the country's/economy's population of 15-year olds. It is also possible to compute the change in the share of 15-year-olds (both those enrolled in school and those not enrolled) who attain higher levels of performance in PISA.

Figure I.2.22 presents the average three-year trend in the median performance of 15-year-olds after accounting for changes over time in the percentage of 15-year-olds that the PISA sample represents (known as Coverage index 3). Only countries where the Coverage index 3 for PISA increased by more than 3 percentage points every three years, on average, are included in this figure (see Chapter 6 for a discussion of Coverage index 3).

The adjusted trend for the median presented in Figure I.2.22 (and for all countries, in Table I.2.4d) neutralises the impact of changes across time in the coverage of the population of 15-year-olds. These changes are related to differences in the selectivity of secondary education. A positive adjusted trend for the median indicates that the quality of education improved for most 15-year-olds: the minimum level of proficiency attained by a majority of 15-year-olds scores has increased over time. By comparing the adjusted trend for the median with the observed (non-adjusted) trend for mean PISA scores over a similar period of time, it is possible to assess the extent to which differences in sample coverage, particularly those related to expansion of secondary education, influence the trends.

Figure I.2.22 ■ **Average three-year trend in median science performance since 2006, after accounting for changes in coverage**

Selected countries



Notes: Statistically significant differences for the average three-year trend are shown in a darker tone (see Annex A3).
The average three-year trend is the average rate of change, per three-year period, between the earliest available measurement in PISA and PISA 2015. For countries and economies with more than one available measurement, the average three-year trend is calculated with a linear regression model.
Only countries where the Coverage index 3 for PISA increased by more than 3% every three years, on average since 2006, are included in this figure.
Countries are ranked in descending order of the average three-year trend in median science performance, after accounting for changes in coverage.
Source: OECD, PISA 2015 Database, Tables I.2.4a and I.2.4d.
StatLink http://dx.doi.org/10.1787/888933432142

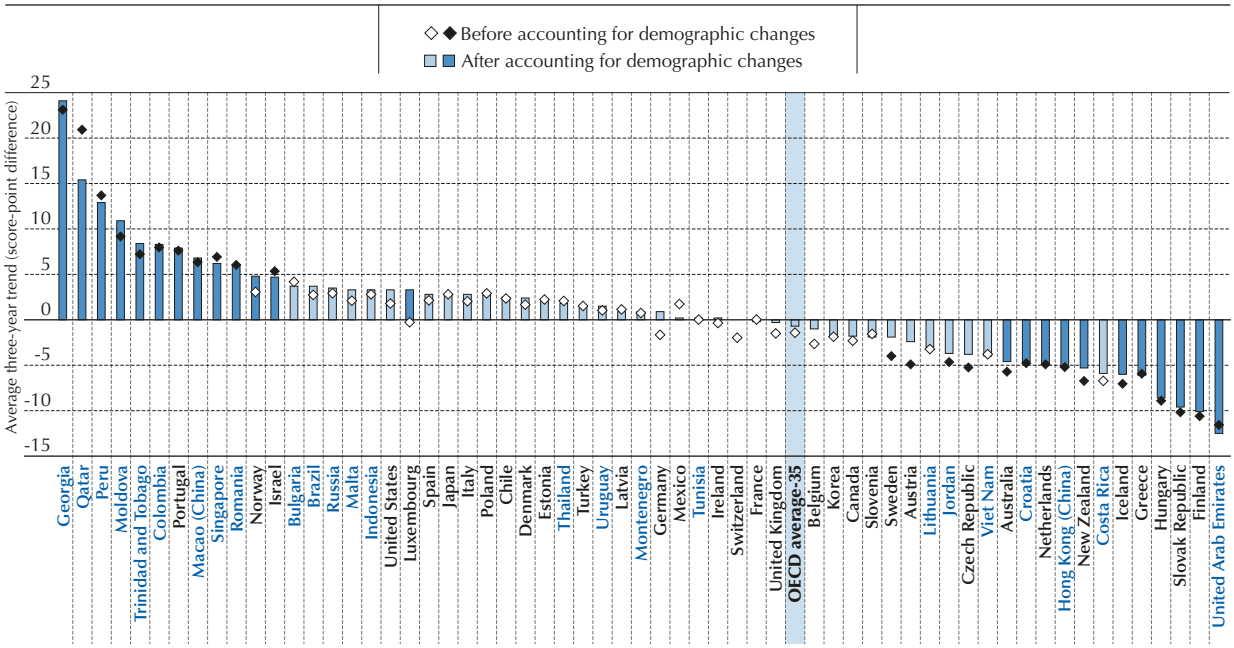
Eleven countries show average increases of at least 3 percentage points every 3 years in the coverage of the PISA sample, indicating that secondary education up to age 15 has become more inclusive in these countries since 2006 (or since the country first participated in PISA). Of these 11 countries and economies, Jordan shows a significant negative mean trend in performance; Brazil, Costa Rica, Indonesia, Russia and Turkey show non-significant trends in performance; and the remaining five (Albania, Colombia, Israel, Portugal and Romania) show a significant positive trend in mean performance (Tables I.2.4a and I.2.4d).

But in all of these countries and economies, the level at which at least 50% of their 15-year-olds perform (the adjusted median) increased significantly between 2006 and 2015 (or since the earliest available assessment), except in Costa Rica, where the increase is not significant. Moreover, the level attained by the 25% best-performing 15-year-olds (adjusted 75th percentile) and the level attained by the 10% best-performing 15-year-olds (adjusted 90th percentile) also rose over the same period in Albania, Brazil, Colombia, Israel, Macao (China), Portugal, Romania and Turkey (in Russia and Indonesia, the increase is significant only at the 75th percentile). This shows that the PISA-participating countries that made their education systems more inclusive over the past decade, as indicated by larger shares of 15-year-olds who are in secondary school, have not done so at the expense of the quality of education for most 15-year-olds – including those students who would have gone to secondary school under the more exclusive conditions of the past (Table I.2.4d).

Average three-year trend in performance, adjusted for demographic changes

In some countries, the demographics of the student population and of the PISA sample have changed considerably across PISA assessments. It is possible to analyse the impact of changes in the immigrant background, age and gender of the student population in each country and economy by contrasting the (unadjusted) changes in mean performance, reported in previous sections, with those that would have been observed had the overall profile of the student population been the same, throughout the period, as that observed in 2015. Adjusted trends in this section provide an estimate of what the performance trend would have been if past PISA samples had the same proportion of immigrant students (first- and second-generation) and the same composition by gender and age (defined in three-month increments) as the target population in 2015.

Figure I.2.23 ■ **Average three-year trend in science performance since 2006, after accounting for demographic changes**



Notes: Statistically significant differences are shown in a darker tone (see Annex A3).

The average three-year trend is the average rate of change, per three-year period, between the earliest available measurement in PISA and PISA 2015. For countries and economies with more than one available measurement, the average three-year trend is calculated with a linear regression model. This model takes into account that Costa Rica, Georgia, Malta and Moldova conducted the PISA 2009 assessment in 2010 as part of PISA 2009+.

The average three-year trend after accounting for demographic changes shows how the performance of a population with the same demographic profile of the PISA 2015 population has changed over time. Demographic characteristics considered are: students' age (in three-month increments), gender, and immigrant background.

Only countries/economies with valid results for PISA 2015 and at least one prior assessment are shown.

Countries and economies are ranked in descending order of the average three-year trend in science performance, after accounting for demographic changes.

Source: OECD, PISA 2015 Database, Tables I.2.4a and I.2.4e.

StatLink http://dx.doi.org/10.1787/888933432151

On average across OECD countries, if the student population in 2006 had the same demographic profile as the population in 2015, the average score in science would have been 496 points. In reality, the average observed score in 2006 was 498 points, and the observed score in 2015 was 493 points. Both the observed and the adjusted trends, therefore, show no significant change, on average, since 2006 (Table I.2.4e).

However, Figure I.2.23 highlights that in Luxembourg, the adjusted trend that neutralises the effects of shifts in the demographic composition of the target population, particularly (in this case) the increase in the percentage of immigrant students, is significant and positive: it corresponds to an increase of about three points every three years since 2006. But the observed trend is flat and not significant: -0.3 points every three years since 2006. This difference in trends before and after accounting for demographic changes means that were it not for these demographic changes, average science performance in Luxembourg would have improved since 2006. Similarly, in Norway, the adjusted trend is significant and positive (+4.8 points per three-year period), but the observed trend is not significant (+3.1 points per three-year period).

Other countries with significantly negative observed trends would not have seen such steep declines in performance were it not for demographic shifts in the composition of the target population. In Austria, the observed trend corresponds to a decline in performance of 4.9 points every three years; but the trend would have been reported as a non-significant decrease of 2.4 points every three years if there had been no concurrent demographic changes. Similarly, in Sweden, the observed trend is negative and significant (-4.0 points), but the adjusted trend is not significant (-2.1 points).

Figure I.2.23 highlights other countries/economies where the demographic shifts in the sample or in the target population influence the observed trends, but where the conclusion about the non-significance of the trend is not affected by these shifts.[12] In Belgium, Germany and Switzerland,[13] in particular, the adjusted trends that account for demographic shifts are more positive, by at least 1.5 points every three years, than the observed trends.

At the opposite end of the spectrum is Qatar, whose positive trends in PISA performance partly reflect favourable shifts in the demographic composition of the target population. In this case, the observed trend shows faster improvement than the adjusted trend that accounts for these shifts; nevertheless, both the observed and the adjusted trends are significant and positive.

Informative as they may be, adjusted trends are merely hypothetical scenarios that help to show the sources of changes in student performance over time. Observed (unadjusted) trends shown in Figure I.2.21 and throughout this chapter summarise the overall evolution of a school system. Comparing observed trends with hypothetical, adjusted trends can, nevertheless, highlight the challenges that countries and economies face in improving students' and schools' science performance.

Comparing mean science performance between 2006 and 2015

At any given point in time, some countries and economies perform similarly. But as time passes and school systems evolve, certain countries and economies improve their performance, pull ahead of the group of countries with which they had shared similar performance levels, and catch up to another group of countries; in other countries and economies, performance falters, and these countries/economies fall behind in rankings relative to other countries. Figure I.2.24 shows, for each country and economy with comparable results in 2006 and 2015, those other countries and economies that performed similarly in science in 2006 but better or worse in 2015.

For example, in 2006, Japan scored at about the same level as Australia, Canada, Korea, the Netherlands and New Zealand, and scored significantly below Finland and Hong Kong (China). But as a result of these countries' negative trends in performance between 2006 and 2015, Japan pulled ahead of all those countries in 2015. In 2006, Portugal scored below France and Spain; but as a result of improvements in Portugal's performance over the same period, by 2015 its mean score in science was higher than that of Spain, and was at the same level as that of France.

Figure I.2.25 shows the relationship between each country's/economy's average science performance in 2006 and the average rate of change between 2006 and 2015. Countries and economies that show the largest improvement throughout the various assessments (top half of the graph) are more likely to be those that performed comparatively poorly in the initial years. The correlation between a country's/economy's earliest comparable science score and the average rate of change is -0.59. This means that 34% of the variation in the rate of change can be explained by a country's/economy's initial score, and that countries with a lower initial score tend to improve at a faster rate.[14]

Although countries that improve the most are more likely to be those that performed relatively poorly in 2006, some countries and economies that scored at or above the average in 2006 also saw improvements in their students' performance over time. Such was the case in Macao (China), which saw improvements in science performance even after its PISA 2006 science scores placed it above the OECD average (results for countries and economies that began their participation in PISA after PISA 2006 are reported in Table I.2.4a).

Other high-performing countries and economies that began their participation in PISA after the 2006 assessment, like Singapore, also show improvements in performance. In addition, there are many countries and economies that performed similarly in 2006 but evolved differently. For instance, Greece and Portugal had scores that were not significantly different from each other's in 2006 (473 points and 474 points, respectively), but in 2015, more than 40 points (the equivalent of more than a year of schooling) separated their mean scores (455 points for Greece and 501 points for Portugal).

Trends in performance among low- and high-achieving students

Changes in a country's or economy's average performance can result from changes at different levels of the performance distribution. For example, for some countries and economies, the average score increased when the share of students scoring at the lowest levels of the science scale shrank because of improved performance among these students. In other countries and economies, improvements in mean scores were largely the result of improvements in performance among the highest-achieving students and an increase in share of students who perform at the highest levels.

Across OECD countries on average, the proportion of students scoring below Level 2 in science increased by 1.5 percentage points between 2006 and 2015 (a non-significant increase), whereas the proportion of students scoring at or above Level 5 decreased by 1.0 percentage point (a non-significant decrease) (Figure I.2.26). Between 2006 and 2015, four countries/economies reduced the share of students who perform below Level 2: Colombia, Macao (China), Portugal and Qatar. While all of these countries reduced the share of low performers, Macao (China), Portugal and Qatar were also able to simultaneously increase the share of students performing at or above Level 5.

Figure I.2.24 [Part 1/4] ■ Multiple comparisons of science performance between 2006 and 2015

Comparison country/economy	Science performance in 2006	Science performance in 2015	Countries/economies with…		
			... similar performance in 2006 and in 2015	... similar performance in 2006, but higher performance in 2015	... similar performancein 2006, but lower performance in 2015
Japan	531	538	Estonia, Chinese Taipei		Canada, Korea, New Zealand, Australia, Netherlands
Estonia	531	534	Japan, Chinese Taipei		Canada, New Zealand, Australia, Netherlands
Chinese Taipei	532	532	Japan, Estonia, Canada		New Zealand, Australia, Netherlands
Finland	563	531			
Macao (China)	511	529			United Kingdom, Germany, Switzerland, Ireland, Belgium, Austria, Czech Republic
Canada	534	528	Chinese Taipei	Japan, Estonia	New Zealand
Hong Kong (China)	542	523			
Korea	522	516	New Zealand, Slovenia, Australia, United Kingdom, Germany, Netherlands	Japan	Czech Republic
New Zealand	530	513	Korea, Australia, Netherlands	Japan, Estonia, Chinese Taipei, Canada	
Slovenia	519	513	Korea, United Kingdom, Germany		Austria, Czech Republic
Australia	527	510	Korea, New Zealand, Netherlands	Japan, Estonia, Chinese Taipei	
United Kingdom	515	509	Korea, Slovenia, Germany, Switzerland, Ireland	Macao (China)	Belgium, Austria, Czech Republic
Germany	516	509	Korea, Slovenia, United Kingdom, Switzerland, Ireland	Macao (China)	Belgium, Austria, Czech Republic
Netherlands	525	509	Korea, New Zealand, Australia	Japan, Estonia, Chinese Taipei	
Switzerland	512	506	United Kingdom, Germany, Ireland, Belgium	Macao (China)	Austria, Czech Republic, Hungary
Ireland	508	503	United Kingdom, Germany, Switzerland, Belgium	Macao (China)	Austria, Sweden, Czech Republic, Hungary
Belgium	510	502	Switzerland, Ireland	Macao (China), United Kingdom, Germany	Austria, Czech Republic, Hungary
Denmark	496	502	Poland, United States		France, Sweden, Spain, Latvia, Hungary, Lithuania, Croatia, Iceland, Slovak Republic
Poland	498	501	Denmark, United States, Sweden		France, Hungary, Croatia
Portugal	474	501			Russia, Italy, Greece
Norway	487	498	United States, France, Spain		Latvia, Russia, Luxembourg, Lithuania, Croatia, Iceland, Slovak Republic
United States	489	496	Denmark, Poland, Norway, France, Spain, Latvia		Russia, Luxembourg, Lithuania, Croatia, Iceland, Slovak Republic
Austria	511	495	Sweden, Czech Republic	Macao (China), Slovenia, United Kingdom, Germany, Switzerland, Ireland, Belgium	Hungary
France	495	495	Norway, United States, Spain, Latvia	Denmark, Poland	Lithuania, Croatia, Iceland, Slovak Republic
Sweden	503	493	Poland, Austria	Ireland, Denmark	Hungary
Czech Republic	513	493	Austria	Macao (China), Korea, Slovenia, United Kingdom, Germany, Switzerland, Ireland, Belgium	
Spain	488	493	Norway, United States, France, Latvia	Denmark	Luxembourg, Lithuania, Croatia, Iceland, Slovak Republic
Latvia	490	490	United States, France, Spain	Denmark, Norway	Luxembourg, Lithuania, Croatia, Iceland, Slovak Republic
Russia	479	487	Luxembourg, Italy	Portugal, Norway, United States	Lithuania, Greece
Luxembourg	486	483	Russia	Norway, United States, Spain, Latvia	Lithuania, Slovak Republic
Italy	475	481	Russia	Portugal	Greece

Note: Only countries and economies with valid results for the PISA 2006 and PISA 2015 assessments are shown.
Countries and economies are ranked in descending order of mean science performance in 2015.
Source: OECD, PISA 2015 Database.
StatLink http://dx.doi.org/10.1787/888933432161

Figure I.2.24 [Part 2/4] ■ Multiple comparisons of science performance between 2006 and 2015

Comparison country/economy	Science performance in 2006	Science performance in 2015	Countries/economies with…			
			... higher performance in 2006, but similar performance in 2015	... higher performance in 2006, but lower performance in 2015	... lower performance in 2006, but similar performance in 2015	... lower performance in 2006, but higher performance in 2015
Japan	531	538		Finland, Hong Kong (China)		
Estonia	531	534	Finland	Hong Kong (China)		
Chinese Taipei	532	532	Finland	Hong Kong (China)	Macao (China)	
Finland	563	531			Estonia, Chinese Taipei, Macao (China), Canada	Japan
Macao (China)	511	529	Chinese Taipei, Finland, Canada, Hong Kong (China)	Korea, New Zealand, Slovenia, Australia, Netherlands		
Canada	534	528	Finland, Hong Kong (China)		Macao (China)	
Hong Kong (China)	542	523			Macao (China), Canada, Korea	Japan, Estonia, Chinese Taipei
Korea	522	516	Hong Kong (China)			Macao (China)
New Zealand	530	513			Slovenia, United Kingdom, Germany	Macao (China)
Slovenia	519	513	New Zealand, Australia, Netherlands			Macao (China)
Australia	527	510			Slovenia, United Kingdom, Germany, Switzerland	Macao (China)
United Kingdom	515	509	New Zealand, Australia, Netherlands			
Germany	516	509	New Zealand, Australia, Netherlands			
Netherlands	525	509			Slovenia, United Kingdom, Germany, Switzerland, Ireland	Macao (China)
Switzerland	512	506	Australia, Netherlands		Denmark, Poland, Portugal, Norway	
Ireland	508	503	Netherlands		Denmark, Poland, Portugal, Norway, United States	
Belgium	510	502			Denmark, Poland, Portugal, Norway, United States	
Denmark	496	502	Switzerland, Ireland, Belgium	Austria, Czech Republic	Portugal, Norway	
Poland	498	501	Switzerland, Ireland, Belgium, Austria	Czech Republic	Portugal, Norway	
Portugal	474	501	Switzerland, Ireland, Belgium, Denmark, Poland, Norway, United States, Austria, France, Sweden	Czech Republic, Spain, Latvia, Luxembourg, Hungary, Lithuania, Croatia, Iceland, Slovak Republic		
Norway	487	498	Switzerland, Ireland, Belgium, Denmark, Poland, Austria, Sweden, Czech Republic	Hungary	Portugal	
United States	489	496	Ireland, Belgium, Austria, Sweden, Czech Republic	Hungary	Portugal	
Austria	511	495			Poland, Portugal, Norway, United States, France, Spain, Latvia	Denmark
France	495	495	Austria, Sweden, Czech Republic	Hungary	Portugal	
Sweden	503	493	Czech Republic		Portugal, Norway, United States, France, Spain, Latvia, Russia	
Czech Republic	513	493			Norway, United States, France, Sweden, Spain, Latvia, Russia	Denmark, Poland, Portugal
Spain	488	493	Austria, Sweden, Czech Republic	Hungary	Russia	Portugal
Latvia	490	490	Austria, Sweden, Czech Republic	Hungary	Russia	Portugal
Russia	479	487	Sweden, Czech Republic, Spain, Latvia	Hungary, Croatia, Iceland, Slovak Republic		
Luxembourg	486	483		Hungary, Croatia, Iceland	Italy	Portugal
Italy	475	481	Luxembourg, Hungary, Lithuania, Croatia	Iceland, Slovak Republic		

Note: Only countries and economies with valid results for the PISA 2006 and PISA 2015 assessments are shown.
Countries and economies are ranked in descending order of mean science performance in 2015.
Source: OECD, PISA 2015 Database.
StatLink http://dx.doi.org/10.1787/888933432161

Figure I.2.24 [Part 3/4] ■ Multiple comparisons of science performance between 2006 and 2015

Comparison country/economy	Science performance in 2006	Science performance in 2015	Countries/economies with…		
			... similar performance in 2006 and in 2015	... similar performance in 2006, but higher performance in 2015	... similar performancein 2006, but lower performance in 2015
Hungary	504	477		Switzerland, Ireland, Belgium, Denmark, Poland, Austria, Sweden	
Lithuania	488	475	Croatia, Iceland	Denmark, Norway, United States, France, Spain, Latvia, Russia, Luxembourg	Slovak Republic
Croatia	493	475	Lithuania, Iceland	Denmark, Poland, Norway, United States, France, Spain, Latvia	Slovak Republic
Iceland	491	473	Lithuania, Croatia	Denmark, Norway, United States, France, Spain, Latvia	Slovak Republic
Israel	454	467			
Slovak Republic	488	461		Denmark, Norway, United States, France, Spain, Latvia, Luxembourg, Lithuania, Croatia, Iceland	
Greece	473	455		Portugal, Russia, Italy	
Chile	438	447	Bulgaria		
Bulgaria	434	446	Chile		Uruguay, Turkey, Jordan
Uruguay	428	435	Romania	Bulgaria	Turkey, Jordan
Romania	418	435	Uruguay, Turkey		Thailand, Mexico, Montenegro, Jordan
Turkey	424	425	Romania, Thailand	Bulgaria, Uruguay	Jordan
Thailand	421	421	Turkey	Romania	Jordan
Qatar	349	418			
Colombia	388	416			Indonesia, Brazil, Tunisia
Mexico	410	416	Montenegro	Romania	
Montenegro	412	411	Mexico	Romania	
Jordan	422	409		Bulgaria, Uruguay, Romania, Turkey, Thailand	
Indonesia	393	403	Brazil	Colombia	Tunisia
Brazil	390	401	Indonesia	Colombia	Tunisia
Tunisia	386	386		Colombia, Indonesia, Brazil	

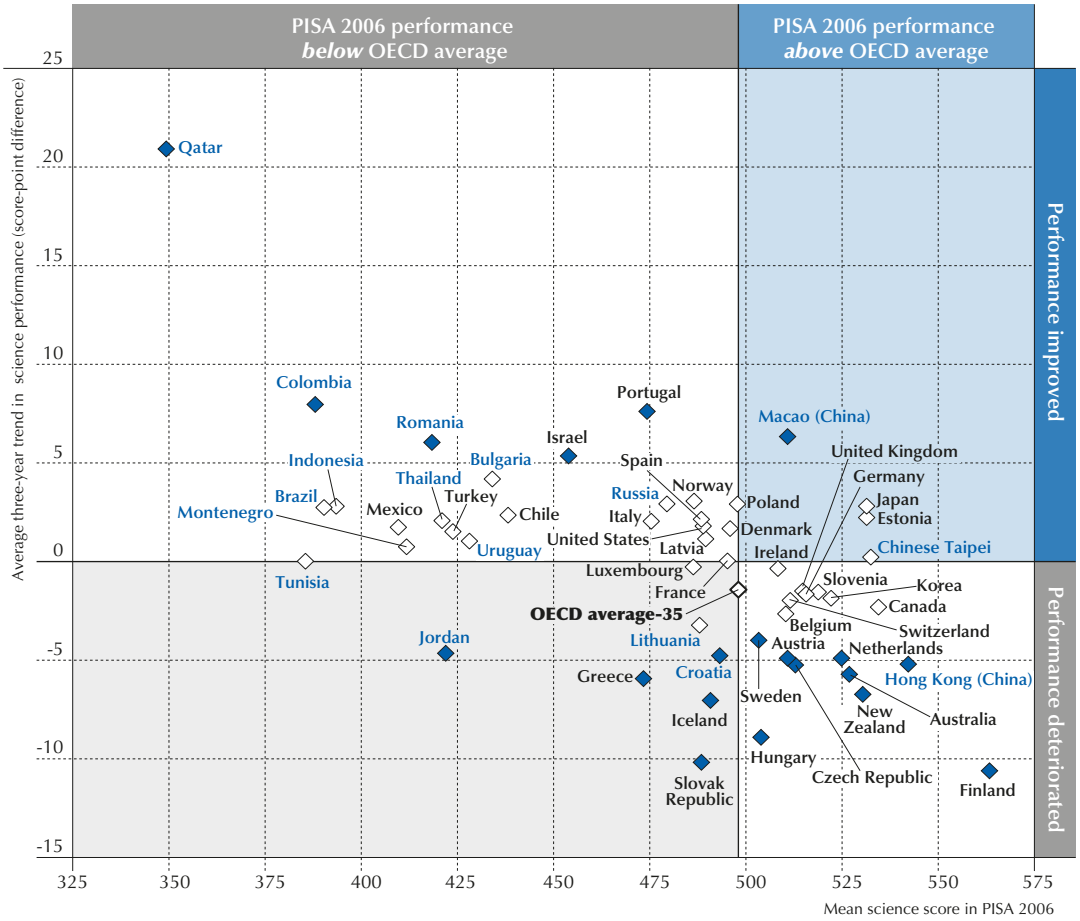
Note: Only countries and economies with valid results for the PISA 2006 and PISA 2015 assessments are shown.
Countries and economies are ranked in descending order of mean science performance in 2015.
Source: OECD, PISA 2015 Database.
StatLink http://dx.doi.org/10.1787/888933432161

Figure I.2.24 [Part 4/4] ■ Multiple comparisons of science performance between 2006 and 2015

Comparison country/economy	Science performance in 2006	Science performance in 2015	Countries/economies with…			
			... higher performance in 2006, but similar performance in 2015	... higher performance in 2006, but lower performance in 2015	... lower performance in 2006, but similar performance in 2015	... lower performance in 2006, but higher performance in 2015
Hungary	504	477			Italy, Lithuania, Croatia, Iceland	Portugal, Norway, United States, France, Spain, Latvia, Russia, Luxembourg
Lithuania	488	475	Hungary		Italy	Portugal
Croatia	493	475	Hungary		Italy	Portugal, Russia, Luxembourg
Iceland	491	473	Hungary		Israel	Portugal, Russia, Luxembourg, Italy
Israel	454	467	Iceland, Slovak Republic	Greece		
Slovak Republic	488	461			Israel, Greece	Portugal, Russia, Italy
Greece	473	455	Slovak Republic		Chile, Bulgaria	Israel
Chile	438	447	Greece			
Bulgaria	434	446	Greece			
Uruguay	428	435				
Romania	418	435				
Turkey	424	425			Qatar	
Thailand	421	421			Qatar, Colombia, Mexico	
Qatar	349	418	Turkey, Thailand, Colombia, Mexico	Montenegro, Jordan, Indonesia, Brazil, Tunisia		
Colombia	388	416	Thailand, Mexico, Montenegro	Jordan	Qatar	
Mexico	410	416	Thailand	Jordan	Qatar, Colombia	
Montenegro	412	411	Jordan		Colombia	Qatar
Jordan	422	409			Montenegro, Indonesia	Qatar, Colombia, Mexico
Indonesia	393	403	Jordan			Qatar
Brazil	390	401				Qatar
Tunisia	386	386				Qatar

Note: Only countries and economies with valid results for the PISA 2006 and PISA 2015 assessments are shown.
Countries and economies are ranked in descending order of mean science performance in 2015.
Source: OECD, PISA 2015 Database.
StatLink http://dx.doi.org/10.1787/888933432161

Figure I.2.25 ■ **Relationship between average three-year trend in science performance and average PISA 2006 science scores**



Notes: Average three-year trends in science that are statistically significant are indicated in a darker tone (see Annex A3).

The average three-year trend is the average rate of change, per three-year period, between the earliest available measurement in PISA and PISA 2015. For countries and economies with more than one available measurement, the average three-year trend is calculated with a linear regression model. This model considers that Costa Rica, Georgia, Malta and Moldova conducted the PISA 2009 assessment in 2010 as part of PISA 2009+.

The correlation between a country's/economy's mean score in 2006 and its average three-year trend is -0.6.

Only countries and economies with available data since 2006 are shown.

Source: OECD, PISA 2015 Database, Table I.2.4a.

StatLink http://dx.doi.org/10.1787/888933432175

Meanwhile, in Australia, the Czech Republic, Finland, Greece, Hungary, New Zealand and the Slovak Republic, the share of students performing at or above Level 5 shrank and, at the same time, the share of students performing below Level 2 grew. In Croatia, the Netherlands and Sweden, the share of low-achieving students increased, but no significant change was observed in the share of top-performing students. And in Austria, Hong Kong (China), Iceland, Ireland, Jordan, Slovenia and the United Kingdom, the share of top performers shrank, but the share of low-achieving students remained stable.

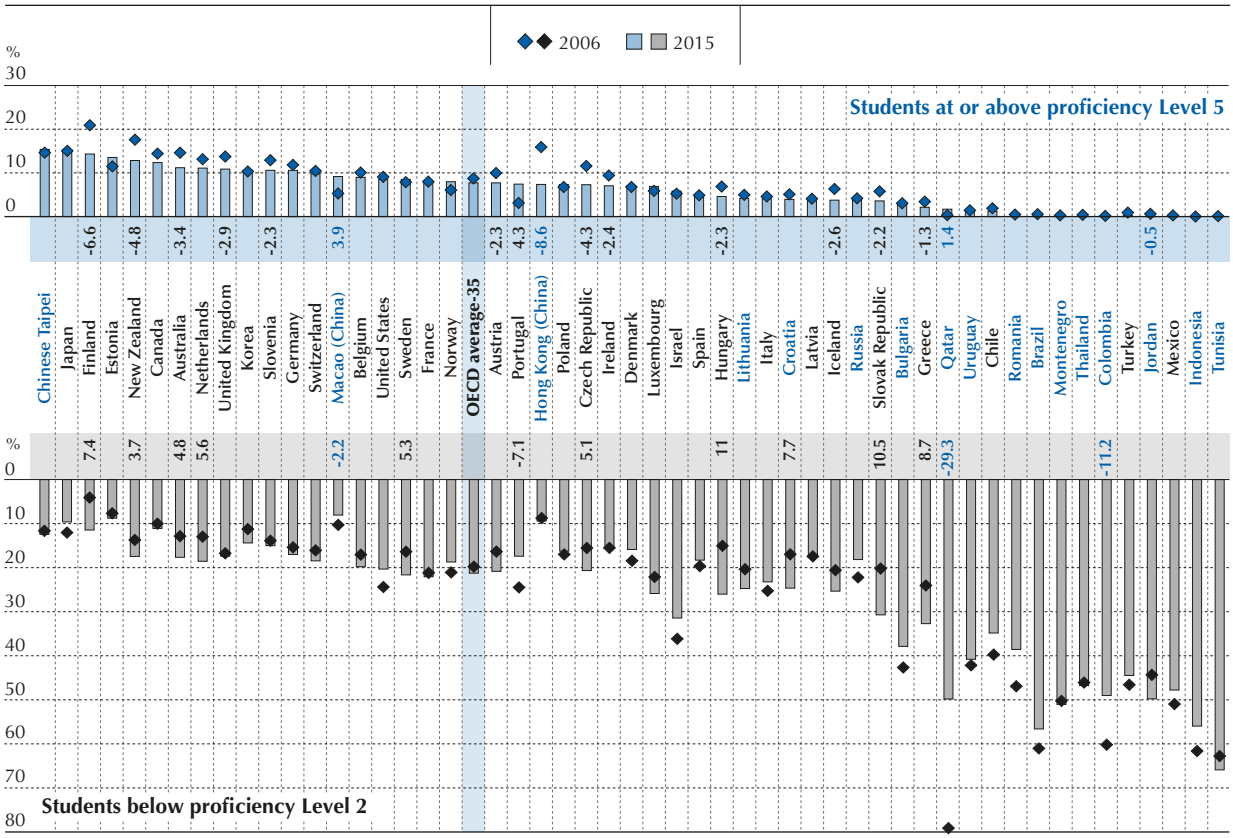
On average across OECD countries, the variation in students' science proficiency remained broadly stable between 2006 and 2015, with similar, non-significant changes across the performance distribution (Tables I.2.4b and I.2.4c).

Between 2006 and 2015, a widening of differences in student performance – measured by the distance between the 10th and the 90th percentile in performance – was observed in Estonia, Finland, Hungary, Korea, Luxembourg, Montenegro, Qatar, the Slovak Republic and Sweden. In Qatar, science performance improved at all levels of the distribution; but the improvement was significantly larger at the top (90th percentile) than at the bottom (10th percentile). In Estonia, Korea, Luxembourg and Montenegro, performance trends at the top (among the highest-achieving students) and at the bottom

(among the lowest-achieving students) show non-significant improvements or declines – but the difference between these trends is significant. In Korea and Sweden, performance remained stable at the top, but declined among the lowest-achieving students. And in Finland, Hungary and the Slovak Republic, performance deteriorated at all levels of proficiency, but more so among the lowest-achieving students (Figure I.2.27 and Table I.2.4c).

Demographic shifts, particularly increases in the immigrant population, sometimes contributed to widening disparities in performance. This is the case in Qatar, where immigrant students typically perform better than non-immigrant students; and in Luxembourg and Sweden, where immigrant students perform worse than non-immigrant students, and their number increased significantly in recent years. In all three countries, however, demographic shifts account for only part of the observed trend. In the remaining countries/economies with widening performance differences, the observed trend at the top and bottom of the performance distribution differs by fewer than 1.5 points from the trends adjusted for shifts in the country's/economy's demographic composition (Table I.2.4f).

Figure I.2.26 ■ **Percentage of low-achieving students and top performers in science in 2006 and 2015**



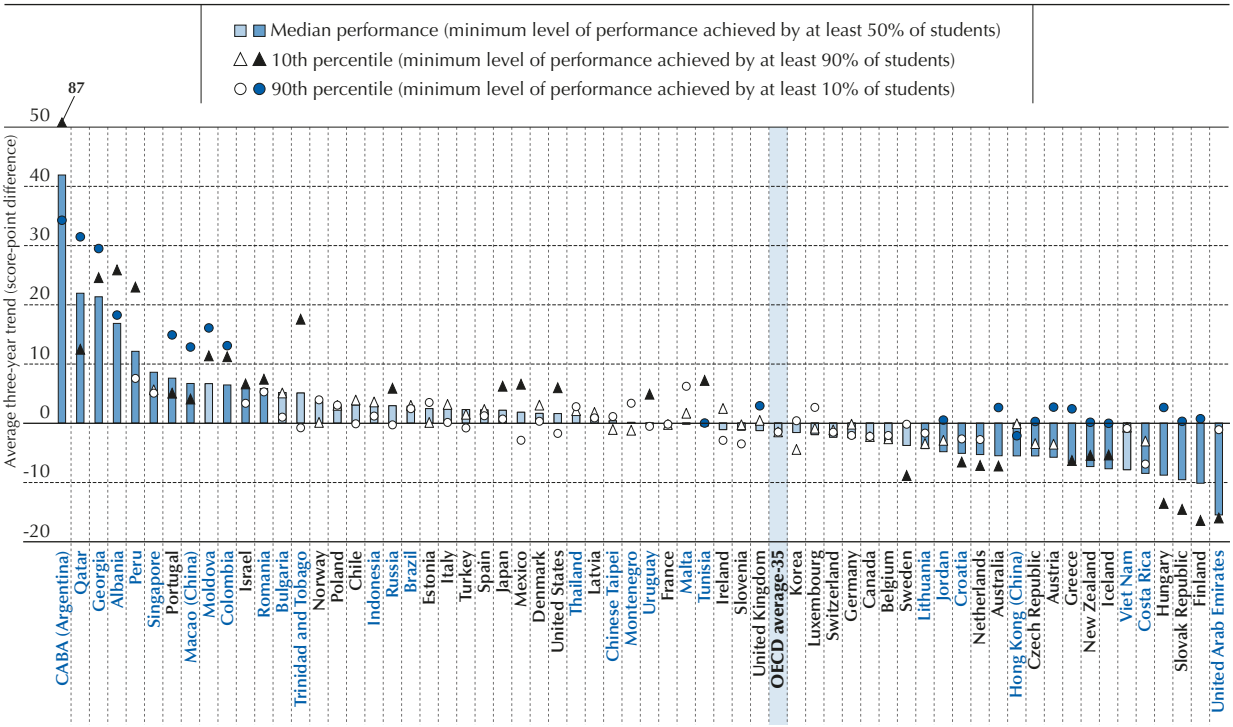
Notes: Only countries/economies that participated in both 2006 and 2015 PISA assessments are shown.
The change between PISA 2006 and PISA 2015 in the share of students performing below Level 2 in science is shown below the country/economy name. The change between PISA 2006 and PISA 2015 in the share of students performing at or above Level 5 in science is shown above the country/economy name.
Only statistically significant changes are shown (see Annex A3).
Countries and economies are ranked in descending order of the percentage of students performing at or above Level 5 in 2015.
Source: OECD, PISA 2015 Database, Table I.2.2a.
StatLink http://dx.doi.org/10.1787/888933432188

Meanwhile, nine other countries and economies (Hong Kong [China], Iceland, Ireland, Mexico, Russia, Tunisia, the United Kingdom, the United States and Uruguay) saw a narrowing of differences in PISA performance. In Mexico, Tunisia, the United States and Uruguay, this reduction reflects improvements among the lowest-performing students, with no significant improvement (and, in the case of Tunisia, a concurrent decline) in performance among the highest-performing students. In Hong Kong (China) and the United Kingdom, performance remained stable at the

10th percentile, but decreased significantly at the top (90th percentile). In Ireland and Russia, neither the positive trend among the lowest-performing students nor the negative trend among the highest-performing students is significant; but the difference between the two trends is significant, and signals a shrinking gap between the top and the bottom. In Iceland, the trend is negative both at the 90th percentile and at the 10th percentile, but more so at the bottom (10th percentile) (Figure I.2.27; Tables I.2.4c and I.2.4f).

Figure I.2.27 ■ **Trends in science performance among high and low achievers**

Average three-year trends in science since 2006



Notes: Statistically significant differences are marked in a darker tone (see Annex A3).

The average three-year trend is the average rate of change, per three-year period, between the earliest available measurement in PISA and PISA 2015. For countries and economies with more than one available measurement, the average three-year trend is calculated with a linear regression model. This model takes into account that Costa Rica, Georgia, Malta and Moldova conducted the PISA 2009 assessment in 2010 as part of PISA 2009+.

Only countries/economies with valid results for PISA 2015 and at least one prior assessment are shown.

Countries and economies are ranked in descending order of the median average three-year trend in science performance.

Source: OECD, PISA 2015 Database, Table I.2.4b.

StatLink http://dx.doi.org/10.1787/888933432199

STUDENTS' PERFORMANCE IN DIFFERENT AREAS OF SCIENCE

In general, scores on any section of the PISA science test are highly correlated with the overall science score. Students who perform well on items classified in one framework category tend to perform well in the other areas of science too. However, at the country level, there is some variation in performance across different subscales. This variation could reflect differences in emphasis in the country's/economy's curriculum. Within the broad domain of science, countries tend to have strong points, where they perform clearly above other countries with otherwise similar performance, and weak points, where they perform worse than countries with similar performance in the remaining areas. This section analyses country's/economy's strong and weak points by looking at differences in mean performance across the PISA science subscales.[15]

Because the science test used in the countries that conducted the PISA 2015 assessment on paper includes only a sample of all science questions, it is not possible to compute subscale scores for these countries with the same reliability as for countries that conducted the PISA 2015 test on computer. For this reason, only countries that used the computer-based science test are included in the following figures and discussion.

Figure I.2.28 ■ Comparing countries and economies on the different science competency subscales

	Mean performance in science (overall science scale)	Mean performance on each science competency subscale			Relative strengths in science: Mean performance on the science competency subscale…[1]		
		Explain phenomena scientifically	Evaluate and design scientific enquiry	Interpret data and evidence scientifically	… explain phenomena scientifically (ep) is higher than on…	… evaluate and design scientific enquiry (ed) is higher than on …	… interpret data and evidence scientifically (id) is higher than on…
Singapore	556	553	560	556		ep id	
Japan	538	539	536	541			ed
Estonia	534	533	535	537			
Chinese Taipei	532	536	525	533	ed		ed
Finland	531	534	529	529	id		
Macao (China)	529	528	525	532			ep ed
Canada	528	530	530	525	id	id	
Hong Kong (China)	523	524	524	521			
B-S-J-G (China)	518	520	517	516			
Korea	516	510	515	523		ep	ep ed
New Zealand	513	511	517	512		ep id	
Slovenia	513	515	511	512	ed		
Australia	510	510	512	508			
United Kingdom	509	509	508	509			
Germany	509	511	506	509	ed		ed
Netherlands	509	509	511	506	id	id	
Switzerland	506	505	507	506			
Ireland	503	505	500	500	ed id		
Belgium	502	499	507	503		ep id	ep
Denmark	502	502	504	500			
Poland	501	501	502	501			
Portugal	501	498	502	503		ep	ep
Norway	498	502	493	498	ed id		ed
United States	496	492	503	497		ep id	ep
Austria	495	499	488	493	ed id		ed
France	495	488	498	501		ep	ep
Sweden	493	498	491	490	ed id		
OECD average	**493**	**493**	**493**	**493**			ed
Czech Republic	493	496	486	493	ed		ed
Spain	493	494	489	493	ed		ed
Latvia	490	488	489	494			ep ed
Russia	487	486	484	489			ed
Luxembourg	483	482	479	486	ed		ep ed
Italy	481	481	477	482			ed
Hungary	477	478	474	476			
Lithuania	475	478	478	471	id	id	
Croatia	475	476	473	476			
Iceland	473	468	476	478		ep	ep
Israel	467	463	471	467		ep id	ep
Slovak Republic	461	464	457	459	ed id		
Greece	455	454	453	454			
Chile	447	446	443	447	ed		
Bulgaria	446	449	440	445	ed id		
United Arab Emirates	437	437	431	437	ed		ed
Uruguay	435	434	433	436			
Cyprus*	433	432	430	434			ed
Turkey	425	426	428	423		id	
Thailand	421	419	423	422			
Costa Rica	420	420	422	415	id	id	
Qatar	418	417	414	418			ed
Colombia	416	412	420	416		ep id	ep
Mexico	416	414	415	415			
Montenegro	411	411	408	410			
Brazil	401	403	398	398	id		
Peru	397	392	399	398		ep	ep
Tunisia	386	385	379	390	ed		ep ed
Dominican Republic	332	332	324	330	ed		ed

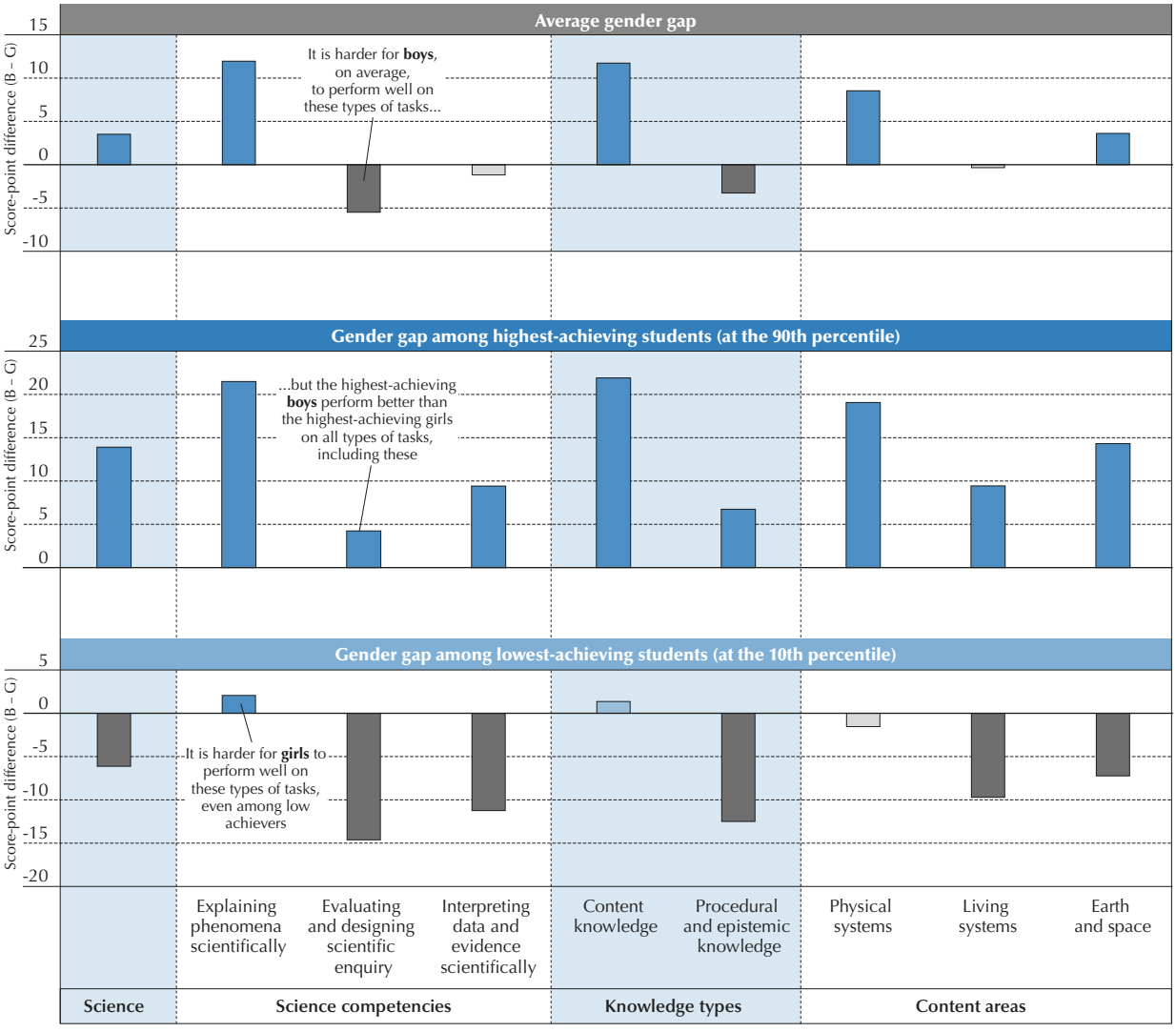
* See note 1 under Figure I.2.13.
1. Relative strengths are highlighted in a darker tone; empty cells indicate cases where the subscale score is not significantly higher compared to other subscales, including cases in which it is lower. Competency subscales are indicated by the following abbreviations: ep – explain phenomena scientifically; ed – evaluate and design scientific enquiry; id - interpret data and evidence scientifically.
Note: Only countries and economies where PISA 2015 was delivered on computers are shown.
Countries and economies are ranked in descending order of mean science performance.
Source: OECD, PISA 2015 Database, Table I.2.13.
StatLink http://dx.doi.org/10.1787/888933432201

Relative strengths and weaknesses of countries/economies in science competency subscales

As discussed above, each item in the PISA 2015 science test was assigned to one of the competency categories, even if solving an item often involved more than one of these competencies. Almost half of all items required that students mainly explain phenomena scientifically; about 30% required them to interpret data and evidence scientifically; and the remaining quarter emphasised the capacity to evaluate and design scientific enquiry. Sometimes, within the same unit, the different items emphasised, in turns, different competencies. Such is the case, for instance, in the released unit *BIRD MIGRATION* (see Annex C1). After a question that asks students to explain phenomena scientifically, in the second question, students must evaluate and design scientific enquiry, and in the last question, they must interpret data and evidence scientifically.

Figure I.2.29 ■ **Boys' and girls' strengths and weaknesses in science**

Score-point difference between boys and girls, OECD average



Notes: All gender differences are statistically significant among the highest-achieving students. Gender differences among average and the lowest-achieving students that are statistically significant are marked in a darker tone (see Annex A3).

Gender differences in favour of girls are shown in grey.

Source: OECD, PISA 2015 Database, Tables I.2.7, I.2.16d, I.2.17d, I.2.18d, I.2.19d, I.2.20d, I.2.21d, I.2.22d and I.2.23d.

StatLink http://dx.doi.org/10.1787/888933432213

Figure I.2.28 shows the country/economy mean for the overall science scale and for each of the competency subscales. It also includes an indication of which differences among the subscale means are significant, through which a country's strengths and weaknesses can be inferred. For instance, while Singapore is the top-performing country in science and in each of the three scientific competencies, it is relatively stronger in students' capacity to evaluate and design scientific enquiry, where the mean performance of students lies clearly above the country's mean performance in the other two competencies (explaining phenomena scientifically and interpreting data and evidence scientifically).

In contrast, students in Chinese Taipei, which appears fourth in the list, are relatively stronger in explaining phenomena scientifically and in interpreting data and evidence scientifically. Korea performs strongest in interpreting data and evidence scientifically, followed by evaluating and designing scientific enquiry, and is comparatively weaker in explaining phenomena scientifically.

Among the remaining countries/economies, Belgium, Israel and the United States stand out for their strong performance in evaluating and designing scientific enquiry in comparison with their performance in explaining phenomena scientifically. France is also relatively weaker in explaining phenomena scientifically. Its comparative strengths are in both evaluating and designing scientific enquiry, and interpreting data and evidence scientifically.

A closer look at gender differences in performance across the different types of science tasks reveals that, in most countries, girls lag behind boys in explaining phenomena scientifically (by 12 score points, on average across OECD countries) (Table I.2.16d). Boys' strength in science lies in their greater capacity, on average, to recall and apply their knowledge of science, identify or generate explanatory models for a situation, and make predictions based on such models. At the same time, boys and girls perform at similar levels when they are asked to interpret data and evidence scientifically (Table I.2.18d). In most countries, girls' relative strength lies in their competency in evaluating and designing scientific enquiry (Table I.2.17d) (Figure I.2.29).

Relative strengths and weaknesses of countries/economies in science knowledge subscales

Science literacy requires an understanding of the major facts, concepts and explanatory theories that form the basis of scientific knowledge. Such understanding encompasses both knowledge of the natural world and of technological artefacts (content knowledge), knowledge of how such ideas are produced (procedural knowledge), and an understanding of the underlying rationale for these procedures and the justifications for using them (epistemic knowledge).

While all items in the PISA 2015 science test were assigned to one of those three knowledge categories, for the purposes of deriving subscales, the latter two categories were combined in the "procedural and epistemic knowledge" subscale. Indeed, there were too few "epistemic knowledge" tasks to support a separate subscale with desirable properties. Approximately half of all the assessment items mainly tested students' content knowledge. Three-quarters of the remaining items assessed procedural knowledge, and the other items (or one-tenth of all science items) aimed to assess students' epistemic knowledge.

Figure I.2.30 shows the country/economy mean for the overall science scale and for the two science knowledge subscales. A dark highlight on the right side of the figure indicates when one of the subscale mean scores is significantly higher than the other. For example, among countries performing close to the OECD average, France and the United States are relatively stronger in their students' capacity to solve questions relating to procedural and epistemic knowledge, whereas Austria, the Czech Republic, Norway and Sweden are relatively stronger in their students' capacity to solve questions relating to content knowledge. Despite these differences on the knowledge subscales, however, the mean scores of these four countries on the overall science scale are not statistically different from each other.

Gender differences in science performance, in favour of boys, are more pronounced when students respond to questions that require content knowledge than when the questions are about procedural or epistemic knowledge (Figure I.2.29). On average across OECD countries, the difference between boys' and girls' scores in science is only 4 points (Table I.2.7); but boys score 12 points higher than girls, on average, on the content knowledge subscale (Table I.2.19d), and girls score 3 points higher than boys on the procedural and epistemic knowledge subscale (Table I.2.20d). This may suggest that, compared with boys, girls are more interested in knowing how scientists enquire and build scientific theories, while boys are relatively more interested in the explanations of natural and technological phenomena that science provides.

Figure I.2.30 ■ Comparing countries and economies on the different science knowledge subscales

	Mean performance in science (overall science scale)	Mean performance on each science knowledge subscale		Relative strengths in science: mean performance on the science knowledge subscale…[1]	
		Content knowledge	Procedural and epistemic knowledge	… content knowledge (co) is higher than on…	… procedural and epistemic knowledge (pe) is higher than on…
Singapore	556	553	558		co
Japan	538	539	538		
Estonia	534	534	535		
Chinese Taipei	532	538	528	pe	
Finland	531	534	528	pe	
Macao (China)	529	527	531		co
Canada	528	528	528		
Hong Kong (China)	523	526	521	pe	
B-S-J-G (China)	518	520	516	pe	
Korea	516	513	519		co
New Zealand	513	512	514		
Slovenia	513	515	512	pe	
Australia	510	508	511		
United Kingdom	509	508	510		
Germany	509	512	507	pe	
Netherlands	509	507	509		
Switzerland	506	506	505		
Ireland	503	504	501	pe	
Belgium	502	498	506		co
Denmark	502	502	502		
Poland	501	502	501		
Portugal	501	500	502		
Norway	498	502	496	pe	
United States	496	490	501		co
Austria	495	501	490	pe	
France	495	489	499		co
Sweden	493	498	491	pe	
OECD average	**493**	**493**	**493**		
Czech Republic	493	499	488	pe	
Spain	493	494	492		
Latvia	490	489	492		co
Russia	487	488	485		
Luxembourg	483	483	482		
Italy	481	483	479	pe	
Hungary	477	480	474	pe	
Lithuania	475	478	474	pe	
Croatia	475	476	475		
Iceland	473	468	477		co
Israel	467	462	470		co
Slovak Republic	461	463	458	pe	
Greece	455	455	454		
Chile	447	448	446		
Bulgaria	446	447	445		
United Arab Emirates	437	437	435		
Uruguay	435	434	436		
Cyprus*	433	430	434		co
Turkey	425	425	425		
Thailand	421	420	422		
Costa Rica	420	421	417	pe	
Qatar	418	416	418		
Colombia	416	413	417		co
Mexico	416	414	416		
Montenegro	411	409	411		
Brazil	401	400	401		
Peru	397	392	399		co
Tunisia	386	386	386		
Dominican Republic	332	331	330		

* See note 1 under Figure I.2.13.
1. Relative strengths are highlighted in a darker tone; empty cells indicate cases where the subscale score is not significantly higher compared to other subscales, including cases in which it is lower. Knowledge subscales are indicated by the following abbreviations: co - content knowledge; pe - procedural and epistemic knowledge.
Note: Only countries and economies where PISA 2015 was delivered on computers are shown.
Countries and economies are ranked in descending order of mean science performance.
Source: OECD, PISA 2015 Database, Table I.2.14.
StatLink http://dx.doi.org/10.1787/888933432228

Relative strengths and weaknesses of countries/economies in science content subscales

The content for the PISA 2015 assessment of science came from topics in the major fields of physics, chemistry, biology, and earth and space science. In order to ensure a balanced representation of different content domains, all items were classified into one of three content areas:

- the "physical" systems content area, comprising all items that require, for example, knowledge of the structure and properties of matter, including its chemical properties, chemical reactions, motion and forces, magnetic fields, energy and its transformation, and interactions between energy and matter
- the "living" systems content area, comprising all items that require, for example, knowledge of the cell and its structures (e.g. DNA), the concept of an organism, human biology, populations (e.g. species and their evolutionary dynamics), ecosystems and the biosphere
- the "earth and space" systems content area, comprising all items that require, for example, knowledge about the structure of earth systems (e.g. atmosphere), changes in earth systems (e.g. plate tectonics), the earth's history, the solar system, and the history and scale of the universe.

Each content category is represented in about one-third of the units in the PISA 2015 assessment. Items, rather than units, were classified according to content system. The classification describes the content knowledge that is required to answer a particular question, rather than general features of the stimulus material. For instance, within the unit *SUSTAINABLE FISH FARMING,* the first three questions are classified in the "living systems" content category while the last question is classified in the "physical systems" category.

Different countries emphasise different topics in their curricula and, depending on their interests and perhaps on the extent to which they are affected by related phenomena (e.g. earthquakes, air pollution or disease), students may be more or less familiar with particular topics that are related to the three content categories in PISA.

Figure I.2.31 shows the country/economy mean for the overall science scale and for the three science content subscales. A highlight on the right side of the panel indicates score differences between subscales that are statistically significant, and signals, for each country/economy, content areas in which performance is relatively strong compared to other areas.

In general, differences across countries/economies mirror those found on the overall science scale, and mean score differences across subscales amount to only a few points. Many countries performing below the OECD average, however, are relatively stronger in the "living systems" content area. This relative strength compared to the two other content areas is particularly marked in Brazil, Peru and Qatar. In these countries/economies, the mean score is at least eight points higher on the living systems subscale than on each of the two other content subscales.

Gender differences in performance across different content areas are broadly similar to overall gender differences in science, with narrower variations than observed across competency or knowledge subscales (Figure I.2.29). Boys outperform girls by nine points, on average across OECD countries, on the physical systems subscale (Table I.2.21d), and by four points on the earth and space systems subscale (Table I.2.23d). Boys and girls have the same mean performance on the living systems subscale, on average (Table I.2.22d).

STUDENTS' EPISTEMIC BELIEFS ABOUT SCIENCE

Science literacy, as defined in PISA, encompasses not only knowledge of the natural world and of technological artefacts (content knowledge), but also knowledge of how such ideas are produced by scientists, and an understanding of the goal of scientific enquiry and of the nature of scientific claims (procedural and epistemic knowledge) (OECD, 2016b). PISA measured whether students are able to use their knowledge about the means and goals of science in order to interpret scientific claims through test items that are classified in the "epistemic knowledge" category, such as those in the unit *SLOPE-FACE INVESTIGATION*.

Through the background questionnaire, PISA 2015 asked students to answer questions about their personal epistemic beliefs about science, i.e. their beliefs about the nature of knowledge in science and about the validity of scientific methods of enquiry as a source of knowing. Students whose epistemic beliefs are in agreement with current views about the nature of science can be said to value scientific approaches to enquiry.

Figure I.2.31 ■ Comparing countries and economies on the different science content subscales

	Mean performance in science (overall science scale)	Mean performance on each science content subscale			Relative strengths in science: mean performance on the science content subscale…[1]		
		Physical systems	Living systems	Earth and space systems	… physical systems (ph) is higher than on…	… living systems (li) is higher than on …	… earth and space systems (es) is higher than on…
Singapore	556	555	558	554		ph es	
Japan	538	538	538	541			
Estonia	534	535	532	539	li		ph li
Chinese Taipei	532	531	532	534			
Finland	531	534	527	534	li		li
Macao (China)	529	533	524	533	li		li
Canada	528	527	528	529			
Hong Kong (China)	523	523	523	523			
B-S-J-G (China)	518	520	517	516			
Korea	516	517	511	521	li		ph li
New Zealand	513	515	512	513			
Slovenia	513	514	512	514			
Australia	510	511	510	509			
United Kingdom	509	509	509	510			
Germany	509	505	509	512		ph	ph
Netherlands	509	511	503	513	li		li
Switzerland	506	503	506	508			ph
Ireland	503	507	500	502	li es		
Belgium	502	499	503	503		ph	ph
Denmark	502	508	496	505	li		li
Poland	501	503	501	501			
Portugal	501	499	503	500		ph	
Norway	498	503	494	499	li		li
United States	496	494	498	496		ph	
Austria	495	497	492	497	li		li
France	495	492	496	496		ph	ph
Sweden	493	500	488	495	li es		li
OECD average	**493**	**493**	**492**	**494**	li		li
Czech Republic	493	492	493	493			
Spain	493	487	493	496		ph	ph
Latvia	490	490	489	493			ph li
Russia	487	488	483	489	li		li
Luxembourg	483	478	485	483		ph	ph
Italy	481	479	479	485			ph li
Hungary	477	481	473	477	li		
Lithuania	475	478	476	471	es	es	
Croatia	475	472	476	477		ph	ph
Iceland	473	472	476	469	es	ph es	
Israel	467	469	469	457	es	es	
Slovak Republic	461	466	458	458	li es		
Greece	455	452	456	453		ph es	
Chile	447	439	452	446		ph es	ph
Bulgaria	446	445	443	448			li
United Arab Emirates	437	434	438	435			
Uruguay	435	432	438	434		ph	
Cyprus*	433	433	433	430		es	
Turkey	425	429	424	421	li es		
Thailand	421	423	422	416	es	es	
Costa Rica	420	417	420	418		ph	
Qatar	418	415	423	409	es	ph es	
Colombia	416	414	419	411		ph es	
Mexico	416	411	415	419		ph	ph
Montenegro	411	407	413	410		ph	
Brazil	401	396	404	395		ph es	
Peru	397	389	402	393		ph es	
Tunisia	386	379	390	387		ph	ph
Dominican Republic	332	332	332	324	es	es	

* See note 1 under Figure I.2.13.
1. Relative strengths are highlighted in a darker tone; empty cells indicate cases where the subscale score is not significantly higher compared to other subscales, including cases in which it is lower. Content subscales are indicated by the following abbreviations: ph - physical systems; li - living systems; es - earth and space systems.
Note: Only countries and economies where PISA 2015 was delivered on computers are shown.
Countries and economies are ranked in descending order of mean science performance.
Source: OECD, PISA 2015 Database, Table I.2.15.
StatLink http://dx.doi.org/10.1787/888933432235

Epistemic beliefs are individuals' representations about the nature, organisation and source of knowledge, e.g. what counts as "true" and how the validity of an argument can be established (Hofer and Pintrich, 1997). When students seek knowledge and understanding, adopt a questioning approach to all statements, search for data and their meaning, demand verification, respect logic and pay attention to premises, they can be said to have a "scientific attitude" and to support scientific approaches to enquiry. Indeed, these are the features that characterise scientific thinking. Such beliefs and dispositions have been shown to be directly related both to students' ability to acquire new knowledge in science and to their grades in school science (Mason et al., 2012).

Epistemic beliefs change with age, as a result of cognitive development and education (Kuhn, Cheney and Weinstock, 2000). In the domain of science, older students are more likely to believe that scientific knowledge is complex, tentative and evolving, is not the property of omniscient authorities, and can be validated in the light of corroborative evidence (Mason et al. 2012). Beliefs about science as an evolving and constantly changing body of knowledge, and about the need for scientific experiments in justifying scientific knowledge, are also related to students' beliefs about learning – particularly to the belief that ability is an incremental, rather than a fixed, attribute (Chen and Pajares, 2010).

PISA did not measure all epistemic beliefs, but focused on measuring students' beliefs about the validity and limitations of scientific experiments and about the tentative and evolving nature of scientific knowledge. It did so through students' responses ("strongly agree", "agree", "disagree" or "strongly disagree") to the statements: "a good way to know if something is true is to do an experiment"; "ideas in science sometimes change"; "good answers are based on evidence from many different experiments"; "it is good to try experiments more than once to make sure of [your] findings"; "sometimes scientists change their minds about what is true in science"; and "the ideas in science books sometimes change". These statements are related to beliefs that scientific knowledge is tentative (to the extent that students recognise that scientific theories are not absolute truths, but evolve over time) and to beliefs about the validity and limitations of empirical methods of enquiry as a source of knowing.

Average levels of support for scientific approaches to enquiry

On average across OECD countries, 84% of students reported that they agree or strongly agree that a good way to know if something is true is to do an experiment; 81% reported that ideas in science sometimes change; 86% reported that good answers are based on evidence from many different experiments; 85% reported that it is good to try experiments more than once to make sure of [your] findings; 80% reported that sometimes scientists change their minds about what is true in science; and 79% reported that the ideas in science books sometimes change (Figure I.2.32).

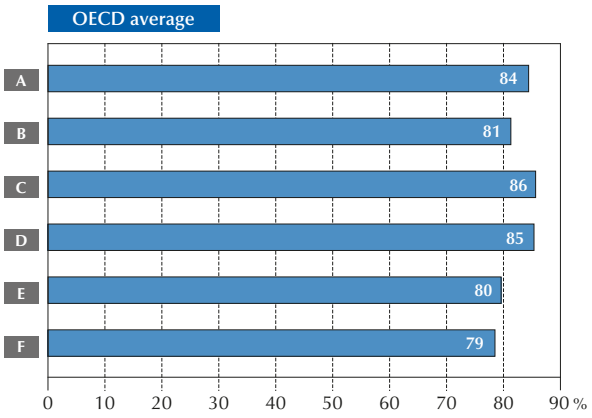
These high percentages reflect broad support for scientific approaches to enquiry, but responses vary markedly among countries and economies. While in Ireland, Singapore and Chinese Taipei more than 93% of students reported that good answers are based on evidence from many different experiments, less than 77% of students in Albania, Algeria, Austria, Montenegro and Turkey agreed with that statement (and more than 23% disagreed) (Table I.2.12a). And while more than nine out of ten students in Australia, Ireland, New Zealand, Portugal, Chinese Taipei, the United Kingdom and the United States agreed that ideas in science sometimes change – reflecting an understanding of science as a changing and evolving body of knowledge – more than one in three students in Austria, Indonesia, Lebanon, Romania and Tunisia disagreed.

Country differences in indices and proportions derived from questionnaire scales must be interpreted with caution, as it is not possible to investigate, with the same rigour applied to test items, whether questionnaire items are equivalent across languages and countries. Because the number of items used to measure self-reported attitudes is limited, a single item whose wording is not understood in the same way across languages may have a disproportionate impact on country/economy rankings on the index derived from these items. Also, a lack of response to the background questionnaire (whether to the entire questionnaire, which is separate from the cognitive test, or to individual questions within the questionnaire) can affect international comparisons. However, the uncertainty about the cross-cultural equivalence of questionnaire scales has less impact on within-country comparisons (e.g. between boys and girls) or on comparisons of associations between questionnaire scales and performance (see Box I.2.4).

Figure I.2.32 ■ Students' epistemic beliefs

Percentage of students who "agree" or "strongly agree" with the following statements

A A good way to know if something is true is to do an experiment
B Ideas in <broad science> sometimes change
C Good answers are based on evidence from many different experiments
D It is good to try experiments more than once to make sure of your findings
E Sometimes <broad science> scientists change their minds about what is true in science
F The ideas in <broad science> science books sometimes change



		A	B	C	D	E	F
OECD	Australia	89	92	92	93	87	86
	Austria	73	63	76	77	67	67
	Belgium	88	82	88	86	82	79
	Canada	89	89	91	92	88	87
	Chile	80	77	81	83	75	71
	Czech Republic	82	79	84	83	81	77
	Denmark	88	85	89	87	89	81
	Estonia	88	85	89	89	83	85
	Finland	84	84	87	87	78	81
	France	88	83	86	84	81	80
	Germany	78	71	79	76	65	66
	Greece	80	70	85	84	75	70
	Hungary	78	71	81	80	68	70
	Iceland	87	88	90	90	87	85
	Ireland	93	92	93	94	82	82
	Israel	86	84	86	86	81	78
	Italy	86	80	84	87	77	76
	Japan	81	82	85	81	76	77
	Korea	86	89	87	88	88	86
	Latvia	81	79	81	77	79	78
	Luxembourg	80	68	80	78	68	68
	Mexico	84	76	83	80	75	77
	Netherlands	85	81	85	85	77	72
	New Zealand	90	91	91	93	86	84
	Norway	84	83	87	85	84	80
	Poland	86	78	85	85	80	83
	Portugal	90	91	91	93	89	90
	Slovak Republic	75	75	78	77	75	73
	Slovenia	89	87	89	90	81	78
	Spain	85	82	87	88	81	81
	Sweden	86	86	87	88	86	84
	Switzerland	81	70	81	80	71	71
	Turkey	73	72	76	76	72	71
	United Kingdom	90	92	91	93	87	87
	United States	90	92	91	92	86	87
Partners	Albania	85	78	75	85	75	89
	Algeria	79	71	75	78	64	65
	CABA (Argentina)	84	85	84	87	80	75
	Brazil	85	84	88	88	82	79
	Bulgaria	81	77	82	80	77	77
	B-S-J-G (China)	89	83	91	87	82	82
	Chinese Taipei	88	94	94	94	93	94
	Colombia	81	77	82	84	75	72
	Costa Rica	79	75	81	83	78	77
	Croatia	89	87	89	85	83	83
	Dominican Republic	78	77	80	80	74	71
	FYROM	78	78	81	84	75	77
	Georgia	86	86	86	86	82	78
	Hong Kong (China)	85	89	90	90	88	86
	Indonesia	92	62	84	90	69	58
	Jordan	75	75	79	81	72	71
	Kosovo	84	80	85	87	74	77
	Lebanon	79	65	81	81	68	67
	Lithuania	81	79	81	79	77	77
	Macao (China)	88	88	91	82	86	85
	Malta	85	86	89	89	76	77
	Moldova	82	83	87	85	80	74
	Montenegro	71	74	77	79	75	75
	Peru	82	79	82	84	76	75
	Qatar	80	78	82	83	77	76
	Romania	76	66	82	79	67	63
	Russia	79	79	83	82	81	78
	Singapore	91	89	94	95	88	87
	Thailand	89	88	89	89	87	87
	Trinidad and Tobago	86	80	87	88	75	75
	Tunisia	78	66	80	82	69	69
	United Arab Emirates	84	82	85	87	80	80
	Uruguay	79	80	80	82	77	77
	Viet Nam	82	82	88	83	78	78

Source: OECD, PISA 2015 Database, Table I.2.12a.

StatLink http://dx.doi.org/10.1787/888933432243

Box I.2.4 Cross-country comparability of questionnaire scales

Most of the indicators of students' science-related beliefs, behaviours and attitudes are based on self-reports. Such measures can suffer from a degree of measurement error, e.g. because students are asked to report their past behaviour retrospectively. Cultural differences in attitudes towards self-enhancement can influence country-level results in students' self-reported beliefs, behaviours and attitudes (Bempechat, Jimenez and Boulay, 2002). The literature consistently shows that response biases, such as social desirability, acquiescence and extreme response choice, are more common in countries with low GDP than in more affluent countries, as they are, within countries, among students from disadvantaged and less-educated families (Buckley, 2009).

In PISA 2015, new scaling methods were introduced to enhance the validity of questionnaire indices, especially for cross-country comparisons. For each item within each scale, an index of item fit was produced for each country-by-language group during the estimation procedure. This fit index provides information about differential item functioning (DIF) across groups and can be used to gauge the overall comparability of scales across countries and language groups.

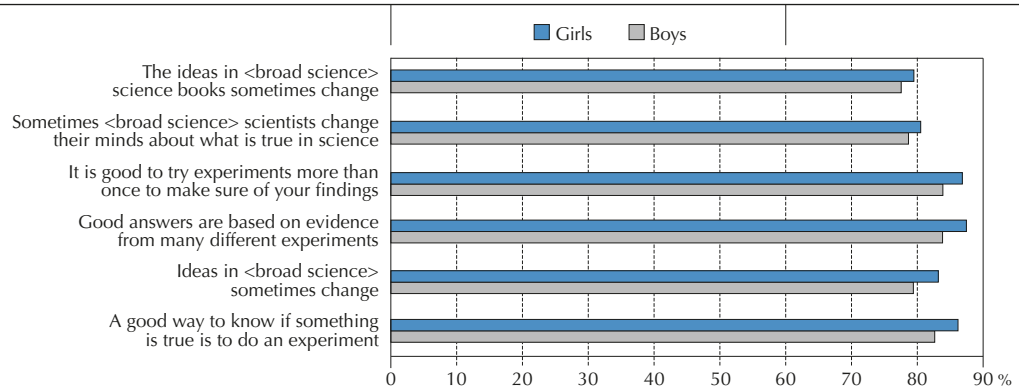
Non-response bias can also affect analyses based on questionnaire items. While statistics based on the science, reading and mathematics proficiency of students are computed on the full PISA sample, student characteristics that are measured through questionnaires are reported as "missing" in the PISA database if the student did not respond to the corresponding question or to the entire questionnaire. The analyses in this report assume that such non-response can be ignored. However, if non-response rates among PISA-participating students are high (e.g., higher than 5% of the sample) and differ significantly across countries, selection bias in the sample used for the analysis may compromise the cross-country comparability of population statistics (such as simple means or correlations with performance). Annex A1 provides for each questionnaire variable used in this volume the percentage of observations for which the information is not missing.

Box I.2.5 Interpreting PISA questionnaire indices

Indices used to characterise students' beliefs and attitudes about science were constructed so that, when they were first developed, the average OECD student would have an index value of zero and about two-thirds of the OECD student population would be between the values of -1 and 1 (i.e. the index has a standard deviation of 1). Therefore, negative values on the index do not imply that students responded negatively to the underlying question. Rather, students with negative values on the index are those who responded less positively than the average response across OECD countries. Likewise, students with positive values on the index are those who responded more positively than the average student in OECD countries (see Annex A1 for a detailed description of how indices were constructed).

Figure I.2.33 ■ **Gender differences in students' epistemic beliefs**

Percentage of students who "agree" or "strongly agree" with the following statements, OECD average

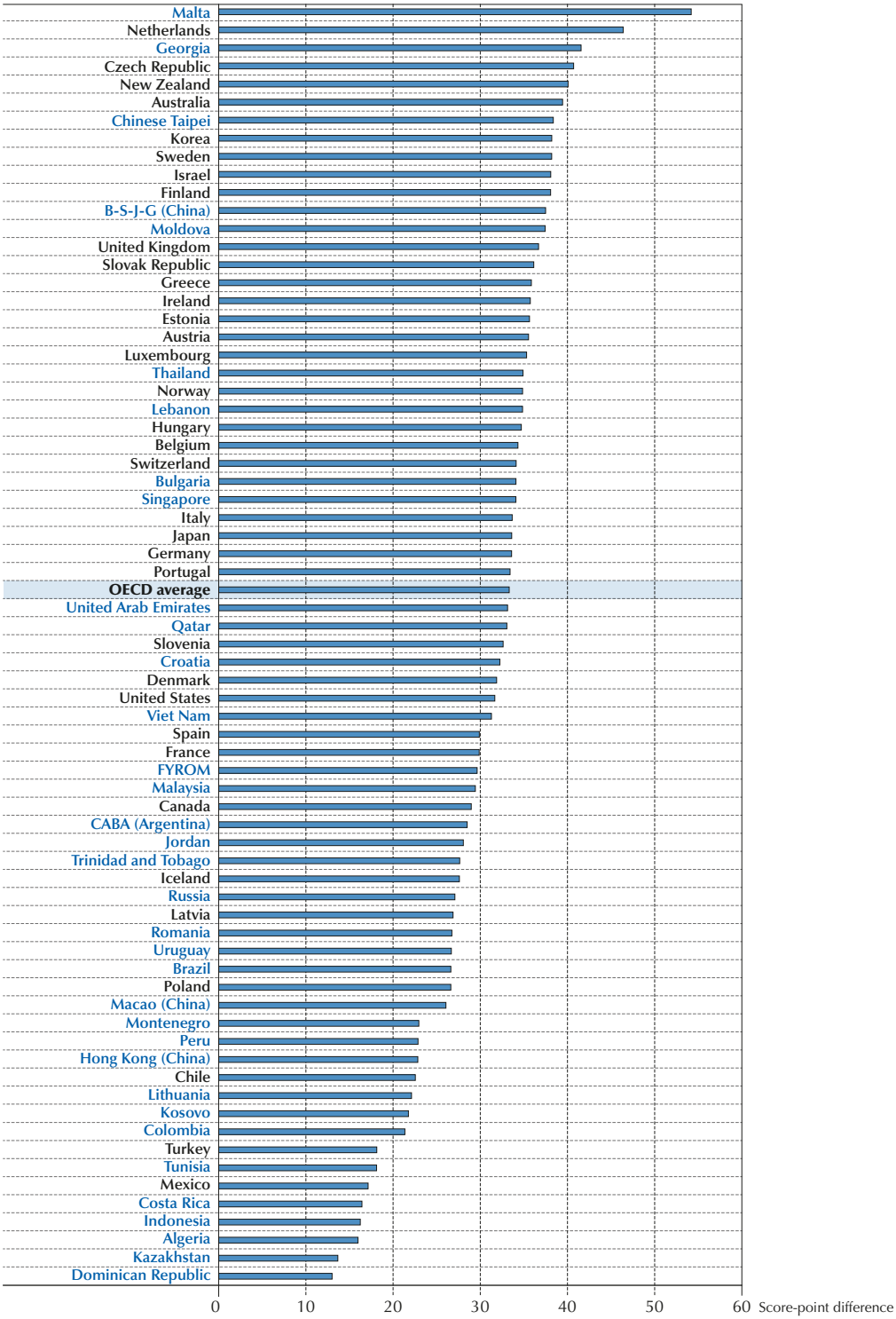


Note: All differences between boys and girls are statistically significant (see Annex A3).
Source: OECD, PISA 2015 Database, Table I.2.12c.

StatLink http://dx.doi.org/10.1787/888933432254

Figure I.2.34 ■ **Relationship between students' belief in scientific approaches to enquiry and science performance**

Score-point difference in science, associated with a one-unit increase on the index of epistemic beliefs



Note: All differences are statistically significant (see Annex A3).

Countries and economies are ranked in descending order of the average score-point difference in science associated with a one-unit increase on the index of epistemic beliefs.

Source: OECD, PISA 2015 Database, Table I.2.12d.

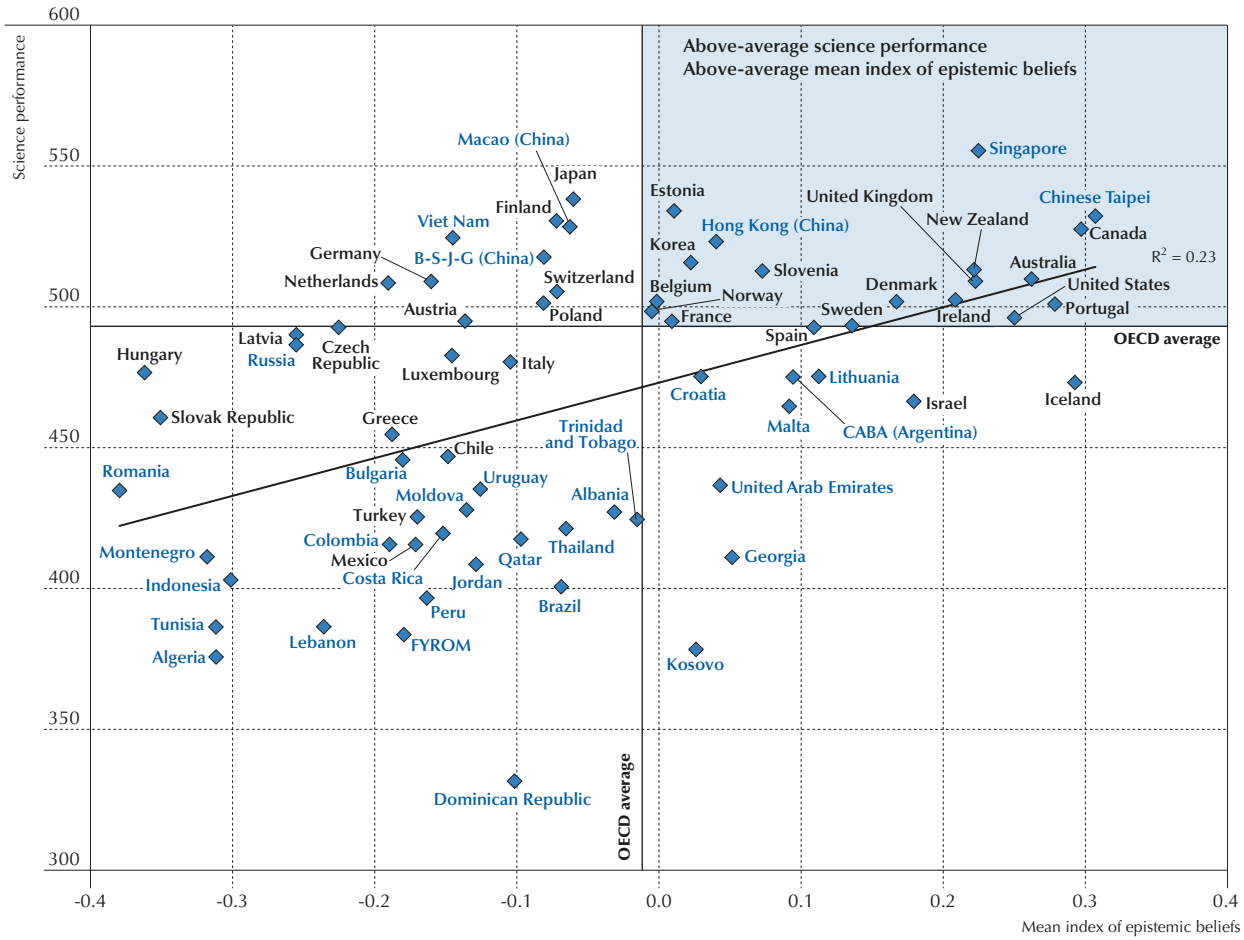
StatLink http://dx.doi.org/10.1787/888933432261

Gender disparities in students' epistemic beliefs are generally small (Figure I.2.33). Where there are differences, the pattern most frequently observed is that of girls reporting more than boys that they support empirical approaches to enquiry as a source of knowing, and that they agree that scientific ideas are tentative and subject to change. The largest such difference between girls and boys is found in Jordan, where 86% of girls reported that a good way to know if something is true is to do an experiment, but only 62% of boys agreed with that statement (Table I.2.12c). Wide differences in favour of girls are also found in FYROM, Georgia, Lithuania and Slovenia.

As Figure I.2.34 indicates, the more strongly students agreed that ideas in science change over time and that experiments provide good ways for establishing whether something is true, the better their performance on the PISA science test, on average. Findings emerging from PISA 2015 cannot be used to establish a direct causal link between personal epistemic beliefs and students' performance on a science test; but PISA shows that the two are closely associated.

The blue bars in Figure I.2.34 denote the estimated difference in science performance that is associated with a difference of one unit on the index of epistemic beliefs about science. This difference corresponds roughly to the difference between a student who "strongly agreed" with the view that a good way to know if something is true is to do an experiment and that it is good to try experiments more than once to make sure of [your] findings, and "agreed" with all other statements; and a student who "agreed" with all statements but one: "disagreeing" with the statement that ideas in science books sometimes change. The former pattern of responses corresponds to an index value of 0.49, half a standard deviation above the OECD average; the latter, to an index value of -0.51.

Figure I.2.35 ■ **System-level association between science performance and students' belief in scientific approaches to enquiry**



Source: OECD, PISA 2015 Database, Tables I.2.3 and I.2.12a.

StatLink http://dx.doi.org/10.1787/888933432270

On average across OECD countries, stronger agreement about the tentative, evolving and cumulative nature of scientific knowledge, and stronger support for empirical approaches to scientific enquiry is associated with higher performance on the PISA science assessment. A one-unit increase on the index corresponds to a 33 score-point difference on the science scale – or about the equivalent of one year of schooling. The fact that all the blue bars represent positive values indicates that in all countries and economies, greater levels of agreement with the questions reflecting students' epistemic beliefs are associated with higher performance. Conversely, higher-performing students tended to "agree" more than lower-performing students with the statements that make up this index.

Differences among students in their epistemic beliefs about science account for about 12% of the variation in students' science performance – similar to the proportion of performance variation that is associated with students' socio-economic status (see Chapter 6). While this association is positive and significant in all countries, the association is markedly weaker in Algeria, Costa Rica, the Dominican Republic, Indonesia, Kazakhstan, Mexico and Tunisia. In these countries/economies, less than 6% of the variation in science performance can be explained by differences in students' science-related epistemic beliefs, and the difference in science performance that is associated with a change of one unit on the index of science epistemic beliefs is less than 20 score points (Table I.2.12b).

At the country/economy level, the mean index of epistemic beliefs has a moderately positive association with science performance, as indicated by a correlation of 0.5. Figure I.2.35 shows that in countries with lower mean performance in science, students were less likely to agree that scientific knowledge is tentative and to support scientific approaches to enquiry. At the same time, among countries with higher mean performance in science, there is a greater variation in students' average beliefs about the nature of scientific knowledge and how such knowledge can be acquired. While this indicates a plausible association that may stem from a cause-effect relationship, the cross-sectional nature of the data and the uncertainty about the cross-cultural equivalence of questionnaire scales does not support firm conclusions about the causal mechanisms at play.

Notes

1. Items that require mainly procedural or epistemic knowledge were also classified depending on the content area or system that provides the context for that knowledge.

2. The results of three countries, however, are not fully comparable, because of issues with sample coverage (Argentina), school response rates (Malaysia), or construct coverage (Kazakhstan); see Annex A4. As a consequence, results for these three countries are not included in most figures.

3. Item difficulty on the PISA scale was defined in PISA 2000 for the purpose of defining proficiency levels as corresponding to a 62% probability of a correct response (Adams and Wu [eds.], 2003, Chapter 16).

4. PISA 2015 science subscales are not directly comparable to PISA 2006 subscales, because they reflect a different way of organising the broad domain of science literacy.

5. In PISA 2006, the mean science score for OECD countries was initially set at 500 points (for 30 OECD countries). Chile, Estonia, Israel and Slovenia acceded to the OECD in 2010. Latvia acceded to the OECD on 1 July 2016. Throughout this report, results for these five countries are included in the OECD average for all cycles of PISA in which they are available. As a result of the inclusion of new countries, the OECD average science score in PISA 2006 is reported as 498 score points.

6. The GDP values represent per capita GDP in 2014 at current prices, adjusted for differences in purchasing power.

7. It should be borne in mind, however, that the number of countries involved in this comparison is small, and that the trend line is therefore strongly affected by the particular characteristics of the countries included in the comparison.

8. Spending per student is approximated by multiplying public and private expenditure on educational institutions per student in 2015 at each level of education by the theoretical duration of education at the respective level, up to the age of 15. Cumulative expenditure for a given country is approximated as follows: let $n(0)$, $n(1)$ and $n(2)$ be the typical number of years spent by a student from the age of 6 up to the age of 15 years in primary, lower secondary and upper secondary education. Let $E(0)$, $E(1)$ and $E(2)$ be the annual expenditure per student in USD converted using purchasing power parities in primary, lower secondary and upper secondary education, respectively. The cumulative expenditure is then calculated by multiplying current annual expenditure E by the typical duration of study n for each level of education i using the following formula:

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

9. The first international comparisons of student proficiency introduced similar assumptions. For instance, the authors of the First International Science Study (FISS) made "the sweeping, but not in general unjustifiable, assumption […] that the members of the population who did not take the test because they had dropped out from secondary school, would have made scores under the 25th percentile, since they had not taken the Science courses" (Comber and Keeves, 1973, pp. 179). In a related exercise, the authors of the First International Mathematics Study (FIMS) compared subgroups of students from each country's total sample that represented the same proportion of the age group as in the country with the lowest coverage rate. For countries with higher coverage rates, only the top part of the distribution was used (Husén 1967, pp. 120-127).

10. For the PISA 2009 assessment, a dispute between teachers' unions and the education minister had led to a boycott of PISA in Austria, which was only withdrawn after the first week of testing. The boycott required the OECD to remove identifiable cases from the Austrian dataset. Although the dataset met the PISA 2009 technical standards after the removal of these cases, the negative atmosphere regarding assessments of education has affected the conditions under which the assessment was administered and could have adversely affected student motivation to respond to the PISA tasks. The comparability of the 2009 data with data from earlier or later PISA assessments cannot, therefore, be ensured for Austria, and 2009 data for Austria have been excluded from trend comparisons.

11. Note by Turkey: In Turkey, students are placed into high schools according to results of national examinations at grade 8. Some 97% of students in the PISA 2015 sample are enrolled in grade 9 or above (21% in grade 9, 73% in grade 10 and 3% in grade 11) and have passed the national examination. The results on the grade 8 exams of students in the PISA 2015 sample who were enrolled in grade 9 or above do not match the expected distribution of results for a representative population of exam-takers. In particular, the top three and the bottom two deciles of exam-takers are under-represented in the PISA sample.

12. The significance of the difference between observed and adjusted trends is not formally tested. Because both trends share a common link error and a perfectly correlated sampling and measurement error (they are estimated on the same samples and data), while each of the estimates is subject to statistical uncertainty, the difference between the two estimates is not subject to these sources of uncertainty.

13. Note by Switzerland: In Switzerland, the increase in the weighted share of students between previous rounds of PISA and PISA 2015 samples is larger than the corresponding shift in the target population according to official statistics.

14. The correlation coefficient exceeds what would be expected under regression to the mean driven solely by (independent) measurement error. In a simulation study, country mean scores were generated using a normal distribution (S.D. = 50 – or about the standard deviation across country mean estimates observed in PISA 2015), along with two independent, noisy measures of these means (with normally distributed noise, S.D. = 3 – or about the typical sampling error for country means in PISA). A Monte Carlo study based on 10 000 simulations shows that the correlation of one of the noisy measures with the difference between the two noisy measures is, on average, -0.04 (95% confidence interval: -0.30 to 0.22).

15. Subscale scores are reported on the same scale as the main science scale. This allows for comparisons across subscales within a particular classification of assessment tasks. Comparisons between subscales related to different classifications – e.g. between a competency subscale and a knowledge-type subscale – or between subscales and the main scale are avoided, however, as it is not possible to correctly estimate from the data the uncertainty associated with such comparisons.

References

Adams, R. and **M. Wu** (eds.) (2003), *Programme for International Student Assessment (PISA): PISA 2000 Technical Report*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264199521-en.

Chen, J.A. and **F. Pajares** (2010), "Implicit theories of ability of grade 6 science students: relation to epistemological beliefs and academic motivation and achievement in science", *Contemporary Educational Psychology*, Vol. 35/1, pp. 75-87, http://dx.doi.org/10.1016/j.cedpsych.2009.10.003.

Comber, L.C. and **J.P. Keeves** (1973), *Science Education in Nineteen Countries,* Wiley & Sons, New York, NY.

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*, pp. 117-149, National Academic Press, Washington, D.C.

Buckley, J. (2009), "Cross-national response styles in international educational assessments: Evidence from PISA 2006", Department of Humanities and Social Sciences in the Professions, Steinhardt School of Culture, Education, and Human Development, New York University, New York, NY.

Davidov, E., P. Schmidt and **J. Billiet** (eds.) (2011), *Cross-Cultural Analysis: Methods and Applications*, Routledge, New York, NY.

Hanushek, E.A. and **L. Woessmann** (2008), "The role of cognitive skills in economic development", *Journal of Economic Literature,* Vol. 46/3, pp. 607-668, http://dx.doi.org/10.1257/jel.46.3.607.

Hofer, B.K. and **P.R. Pintrich** (1997), "The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning", *Review of Educational Research*, Vol. 67/1, pp. 88-140, http://dx.doi.org/10.3102/00346543067001088.

Husén, T. (ed.) (1967), *International Study of Achievement in Mathematics: A Comparison of Twelve Countries* (Vols. I and II), Wiley & Sons, New York, NY.

Janssen, R. (2011), "Using a differential item functioning approach to investigate measurement invariance", in E. Davidov, P. Schmidt and J. Billiet (eds.), *Cross-Cultural Analysis: Methods and Applications*, pp. 415-432, Routledge, New York, NY.

Keskpaik, S. and **F. Salles** (2013), "Les élèves de 15 ans en France selon PISA 2012 en culture mathématique : baisse des performances et augmentation des inégalités depuis 2003", *Note d'information* 13.31, MEN-DEP, Paris, France.

Kuhn, D., R. Cheney and **M. Weinstock** (2000), "The Development of Epistemological Understanding", *Cognitive Development*, Vol. 15/3, pp. 309-328, http://dx.doi.org/10.1016/S0885-2014(00)00030-7.

Mason, L. et al. (2012), "Besides knowledge: A cross-sectional study on the relations between epistemic beliefs, achievement goals, self-beliefs, and achievement in science", *Instructional Science*, Vol. 41/1, pp. 49-79, http://dx.doi.org/10.1007/s11251-012-9210-0.

Meredith, W. (1993), "Measurement invariance, factor analysis and factorial invariance", *Psychometrika*, Vol. 58/4, pp. 525-543, http://dx.doi.org/10.1007/BF02294825.

OECD (forthcoming), *PISA 2015 Technical Report*, PISA, OECD Publishing, Paris.

OECD (2016a), *Low-Performing Students: Why They Fall Behind and How To Help Them Succeed*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264250246-en.

OECD (2016b), "PISA 2015 Science Framework", in *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264255425-en.

OECD (2013), *PISA 2012 Results: What Makes Schools Successful (Volume IV): Resources, Policies and Practices*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264201156-en.

OECD (2012), *Learning beyond Fifteen: Ten Years after PISA*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264172104-en.

OECD (2010), *PISA 2009 Results: What Makes a School Successful?: Resources, Policies and Practices (Volume IV)*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264091559-en.

OECD (2009), *Take the Test: Sample Questions from OECD's PISA Assessments*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264050815-en.

OECD (2007), *PISA 2006: Science Competencies for Tomorrow's World: Volume 1: Analysis*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264040014-en.

OECD, E. Hanushek and **L. Woessmann** (2015), *Universal Basic Skills: What Countries Stand to Gain*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264234833-en.

Osborne, J., S. Simon and **S. Collins** (2003), "Attitudes towards science: A review of the literature and its implications", *International Journal of Science Education,* Vol. 25/9, pp. 1049-1079, http://dx.doi.org/10.1080/0950069032000032199.

Prenzel, M. et al. (eds.) (2006), *PISA 2003: Untersuchungen Zur Kompetenzentwicklung Im Verlauf Eines Schuljahres*, Waxmann Verlag GmbH, Münster, Germany.

Schibeci, R.A. (1984), "Attitudes to science: An update", *Studies in Science Education, Vol.* 11/1, pp. 26-59, http://dx.doi.org/10.1080/03057268408559913.

Spaull, N. and **S. Taylor** (2015), "Access to what? Creating a composite measure of educational quantity and educational quality for 11 African countries", *Comparative Education Review*, Vol. 59/1, pp. 133-165, http://dx.doi.org/10.1086/679295.

Taylor, S. and **N. Spaull** (2015), "Measuring access to learning over a period of increased access to schooling: the case of southern and Eastern Africa since 2000", *International Journal of Educational Development*, Vol. 41, pp. 47-59, http://dx.doi.org/10.1016/j.ijedudev.2014.12.001.

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2004), *Universal primary completion in Latin America: are we really so near the goal? Regional Report on Education-related Millennium Development Goal*, UNESCO Regional Office for Education in Latin America and the Caribbean, Santiago, Chile. http://unesdoc.unesco.org/images/0013/001389/138995eb.pdf.

Woessmann, L. (2016), "The importance of school systems: Evidence from international differences in student achievement", *The Journal of Economic Perspectives*, Vol. 30/3, pp. 3-31, http://dx.doi.org/10.1257/jep.30.3.3.