PISA FOR DEVELOPMENT ASSESSMENT AND ANALYTICAL FRAMEWORK

Draft version, 03 May 2017

FOREWORD

The OECD Programme for International Student Assessment (PISA), created in 1997, represents a commitment by the governments of OECD and partner countries to monitor the outcomes of education systems, in terms of student achievement, within a common, internationally agreed framework. PISA is a collaborative effort, bringing together scientific expertise from the participating countries/economies and steered jointly by their governments on the basis of shared policy interests. Experts from participating countries also serve on working groups that are charged with linking the PISA policy objectives with the best available substantive and technical expertise in the field of internationally comparable assessments. Through involvement in these expert groups, countries ensure that the PISA assessment instruments are internationally valid and take into account the cultural and curricular context of the PISA-participating countries and economies.

Participation in PISA by non-OECD countries is growing and is combined with demand from these countries for innovations that will maximise their benefits from participation in the assessment. PISA for Development (PISA-D) is an initiative that has been developed in response to this demand and in the context of the Education Sustainable Development Goal that was adopted by the international community in 2015 and which emphasises universal access to literacy and numeracy. This pilot project aims to make PISA more accessible and relevant to middle- and low-income countries. It does this by developing and piloting enhanced PISA survey instruments that are more relevant for the contexts found in middle- and low-income countries but which produce scores that are on the same scales as the main PISA assessment. The initiative also includes the development of an approach and methodology for including out-of-school children in the surveys. All of the instruments and approaches piloted in PISA-D will be mainstreamed in PISA from the 2021 edition of the assessment onwards.

This publication presents the guiding principles behind the PISA-D assessment for both the school-based and the out-of-school instruments. Sample tasks are also included. It assembles versions of the PISA assessment frameworks for reading, mathematical and scientific literacy that are based on the PISA 2012 and PISA 2015 frameworks but extends these frameworks to allow for more relevant measurement in a broad range of middle- and low-income countries. Making the measurement more relevant to these countries requires more detail in the description of competencies of the most vulnerable students, those with the lowest levels of performance, which in turn requires including items that will enable the observation of these competencies in greater detail. Yet the relevance of PISA-D, and the aim of mainstreaming the outputs from the initiative in main PISA, also depends on comparability with international PISA results: the instrument therefore allows for students to demonstrate the full range of proficiency levels in PISA.

As with previous cycles of PISA, the PISA-D cognitive frameworks have been reviewed and updated by a network of international experts who have experience with PISA, the relevant domains and the contexts found in middle- and low-income countries. A group of Pearson content experts lead by a chair, together with three experts suggested by the participating countries, reviewed existing versions of the PISA 2012 and 2015 assessment frameworks and prepared Chapters 2, 3 and 4 of this publication. The reading framework was led by Jean-François Rouet, the mathematics framework by Zbigniew Marciniak and the science framework by Jonathan Osborne. Additionally, representatives of the participating countries have provided advice and guidance in the development and extension of the PISA-D framework.

This document also includes the framework for the PISA-D questionnaires, Chapter 5, that were redeveloped for PISA-D. This questionnaires framework was developed by J. Douglas Willms from The Learning Bar Inc. of Canada, with contributions from Lucia Tramonte and Robert Laurie. The questionnaire expert group included Servaas Van der Berg (South Africa, Chair), Andres Peri (Uruguay), Fons van de Vijver (Netherlands), and Wilima Wadhwa (India). Other experts who contributed to the development of the questionnaire framework include Danielle Durepos and Alma Lopez-Garcia.

The publication was prepared by the OECD Secretariat, principally by Michael Ward, Catalina Covacevich and Kelly Makowiecki. A full list of all contributing experts and support staff is included at Annex A. The report is published under the responsibility of the Secretary-General of the OECD.

TABLE OF CONTENTS

Tables

Figures

Boxes

CHAPTER 1. WHAT ARE PISA AND PISA FOR DEVELOPMENT?

“What is important for citizens to know and be able to do?” In response to that question and to the need for cross-nationally comparable evidence on student performance, the Organisation for Economic Co-operation and Development (OECD) launched the Programme for International Student Assessment (PISA) in 1997. PISA assesses the extent to which 15-year-old students, near the end of their compulsory education, have acquired key knowledge and skills that are essential for full participation in modern societies.

The triennial assessment focuses on the core school subjects of reading, mathematics and science. Students’ proficiency in an innovative domain is also assessed. The assessment does not just ascertain whether students can reproduce knowledge; it also examines how well students can extrapolate from what they have learned and can apply that knowledge in unfamiliar settings, both in and outside of school. This approach reflects the fact that modern economies reward individuals not for what they know, but for what they can do with what they know.

Through questionnaires distributed to students, parents, school principals and teachers, PISA also gathers information about students’ home background, their approaches to learning and their learning environments.

In each round of PISA, one of the core domains is tested in detail, so a thorough analysis of achievement in each of the three core areas is presented every nine years and an analysis of trends is offered every three years. Combined with the information gathered through the various questionnaires, the PISA assessment provides three main types of outcomes:

- Basic indicators that provide a baseline profile of the knowledge and skills of students.
- Indicators derived from the questionnaires that show how such skills relate to various demographic, social, economic and educational variables.
- Indicators on trends that show changes in outcome levels and distributions, and in relationships between student-level, school-level and system-level background variables and outcomes.

PISA is an ongoing programme that, over the longer term, will lead to the development of a body of information for monitoring trends in the knowledge and skills of students in various countries as well as in different demographic subgroups of each country. Policy makers around the world use PISA findings to gauge the knowledge and skills of students in their own country/economy in comparison with those in other participating countries/economies, establish benchmarks for improvements in the education provided and/or in learning outcomes, and understand the relative strengths and weaknesses of their own education systems.

The experience of middle-income countries in PISA

Students representing more than 80 countries and economies that together make up over 80% of the world economy have participated in PISA since its launch, including 44 middle-income countries, 27 of which have been recipients of foreign aid. As more and more participants join it has become apparent that the design and implementation models for PISA need to evolve to successfully cater to a larger and more diverse set of countries, including the growing number of middle- and low-income countries who want to participate in the assessment (Lockheed, 2015). In particular, PISA needs to take more account of the marked differences between high- and middle-income countries in education quality and equity and their correlates.

The OECD's analysis of the experience of middle-income countries in PISA has revealed the following three key results that have implications for the further development of the assessment and its framework:

- First, the overall performance of 15-year-old students in all of the middle-income countries participating in PISA, except Viet Nam, is lower than that of students in OECD countries, and varies widely. Performance is also concentrated at the lower levels of the PISA proficiency scales.

- Second, some of the educational inputs as currently measured by PISA are unrelated to differences in performance across schools in the majority of the middle-income countries that participate in PISA. In addition, the measure of economic, social and cultural status currently used by PISA does not adequately capture lower levels of parental education, income and risk factors of poverty that are more frequent in low-income countries. Moreover, it has also become clear that the data captured on the context that surrounds students could be made more relevant, particularly in respect of policies, for middle- and low-income countries.

- Third, out-of-school rates for lower secondary school children are high in many middle- and low-income countries and, in addition, many 15-year-olds in these contexts are also enrolled in grades below those that are eligible for PISA (i.e. below grade 6). The combination of these two exclusion mechanisms result in indices of coverage of the 15-year-old population as low as 50% in some PISA participating countries and limit the comparability of the results of middle-income countries with other countries. It is also the case that PISA runs the risk of reinforcing policies of exclusion in middle-income countries unless the assessment takes concrete steps to incorporate all of the 15-year-olds in a country's population in the survey.

PISA for Development

Building on the experience of middle-income countries in PISA and in an effort to respond to the three results highlighted above, the OECD launched the PISA for Development (PISA-D) initiative in 2014. This is a one-off pilot project spanning six years which aims to make the assessment more accessible and relevant to a wider range of countries. The project is also a contribution to the monitoring of international educational targets related to the Education Sustainable Development Goal (SDG) adopted by the United Nations General Assembly in 2015 as part of the *Agenda for Sustainable Development*. The project has also been informed by analysis of the lessons and experiences from other regional and international large-scale assessments in education in middle- and low-income countries (Cresswell, Schwantner and Waters, 2015). To accomplish its aims, the project sets out to:

- increase the resolution of the PISA tests at the lower end of the student performance distribution

- capture a wider range of social and economic contexts

- incorporate an assessment of out-of-school 14-16-year-olds.

The highly collaborative PISA-D project is being carried out by the OECD, nine participating countries, international contractors, development partners and technical partners.

Eight countries are participating in the school-based implementation of PISA-D: Bhutan, Cambodia, Ecuador, Guatemala, Honduras, Paraguay, Senegal and Zambia. One of the main reasons for their participation is policy makers' wish to understand why students in their countries achieve certain levels of performance. Assessment results will provide these policy makers with data and evidence that can be used to determine what they can do to improve their educational systems and, ultimately, ensure that their students obtain the skills needed to succeed in tomorrow's world and as set out in the Education SDG Framework.

In addition to the school-based component of PISA-D, an out-of-school component is being piloted by six countries – Guatemala, Honduras, Senegal, Paraguay, Panama and Zambia – and focuses on the knowledge, skills and contextual factors of 14-16-year-old out-of-school youth. In PISA-D the definition of out-of-school youth incorporates all those 14-16-year-olds that are not reflected in the school-based survey, including the out-of-school and those who are in school but enrolled at grades 6 or below. This out-of-school component adopts the same framework used for the school-based component of PISA-D, as the description of competencies, particularly at lower levels of performance, will also apply to the out-of-school population. Through the out-of-school assessment, PISA-D will, for the first time, be able to report on what all 15-year-olds in a population know and can do. The analysis of these data should yield valuable insights for governments in middle- and low-income countries in particular about the effectiveness of their education systems, and about the success of policies that aim to ensure inclusive and equitable quality education and learning opportunities for all. It will also serve to reinforce these policies of inclusion and contribute to the monitoring and achievement of the Education SDG with its emphasis on leaving no one behind.

Box 1.1 The out-of-school component

Across many middle- and low-income countries, relatively large proportions of 15-year-olds are not enrolled in school or are enrolled in school in grades below PISA's target grades (grade 7 and above) and are therefore excluded from the PISA sample. In the PISA-D participating countries, between 10 and 50% of youth are in this situation. The PISA-D out-of-school component is establishing methods and approaches to include out-of-school youth aged 14 to 16 and also 14-16-year-old students that are in grades 6 or below in the assessment. The sample range was expanded from 15-year-olds to 14- and 16-year-olds following the recommendations of Carr-Hill (2015), who highlighted the challenge of locating a single year age group in a household survey in middle- and low-income countries. The range of educational experiences in this out-of-school population is expected to vary substantially, from children with no experience in formal education to those who have recently left school or who are still in school but in grades 6 or below.

The PISA-D instruments, once piloted and finalised, will be available for use in future PISA cycles (from PISA 2021 onwards) and will allow middle- and low-income countries to participate in PISA more meaningfully. The enhanced instruments will also support global measures of reading and mathematical skills as part of the Education SDG agenda, strengthening PISA's potential to provide a global metric for measuring progress towards the Education SDG targets and indicators.

The PISA-D framework maintains the concept of competency that was adopted by the PISA Governing Board as part of the long-term strategy for PISA in 2013, which seeks to go beyond the reproduction of subject-matter knowledge and focuses on the capacity of students to extrapolate from what they know and apply their knowledge. Furthermore, the PISA-D framework maintains the same design parameters that have guided all assessments from PISA 2000.

This publication presents the theory underlying the PISA-D assessment, which has been developed in the context of PISA. It includes frameworks for assessing the three core subjects – reading, mathematics and science (Chapters 2, 3 and 4, respectively), that build on the PISA 2012 and 2015 frameworks. The chapters outline the cognitive processes or competencies involved in the tasks of each testing domain, and the area of knowledge and contexts or situations in which these cognitive processes are applied. They also discuss how each domain is assessed. Chapter 5 explains the theory underlying the context questionnaires distributed to students, school principals and teachers, and also the ones answered by the out-of-school youth, their parents (or the person most knowledgeable about the youth), and the interviewer.

What makes PISA-D different to PISA

While PISA-D is being implemented within the overall PISA framework and in accordance with PISA's technical standards and usual practices, it includes new features and enhancements to make the assessment more accessible and relevant to middle- and low-income countries. These features and enhancements include:

- An equal treatment of the three major domains tested: reading, mathematics and science – unlike PISA where in each cycle one of the domains is given a particular focus.

- Targeted test instruments that cover a wider range of performance at the lower levels of proficiency while still providing scores that cover the whole of the PISA framework and are comparable to the main PISA results – unlike PISA where the tests are not targeted on particular levels of performance.

- Modified test instruments and questionnaires that have a reduced reading burden in recognition of the lower levels of reading literacy capacity in middle- and low-income countries.

- Contextual questionnaires that have at their core items from PISA to facilitate international comparisons, but also include several distinct PISA-D items that are more relevant to middle- and low-income countries. These new items also respond to the policy priorities of the countries participating in PISA-D.

- An assessment of the out-of-school population: PISA assesses 15-year-olds that are in school in grades 7 or above. PISA-D assesses this same population, but also has an out-of-school module aimed at 14-16-years-olds who are not in school, or are in school but in grades 6 or below. The inclusion of out-of-school youth in the survey makes PISA-D unique in the landscape of international large-scale assessments. The project explores methodologies and data-collection tools regarding out-of-school youth, i) in terms of their skills, competencies and non-cognitive attributes, and ii) in terms of obtaining better actionable data on the characteristics of these children, the reasons for their not being in school and on the magnitudes and forms of exclusion and disparities.

Another feature unique to PISA-D is the learning and capacity-building opportunities that are built into each phase of project implementation. In preparing to implement the assessment, PISA-D countries undergo a capacity needs analysis based on PISA's technical standards and devise a capacity building plan that is also relevant for strengthening their national assessment systems. The PISA-D countries are also assisted by the OECD in the preparation of a project implementation plan that guides their implementation of the survey and ensures the necessary human and financial resources are in place. While PISA countries have not benefitted from similar support, the PISA-D project serves as the basis for developing a model of support within the core PISA survey which can be offered more widely to all participating countries from the 2021 cycle onwards.

Unlike PISA, PISA-D results will be published in national reports produced by the countries in collaboration with the OECD. As part of the report production process the OECD and its contractors will provide inputs to the countries to strengthen their capacities for data analysis, interpretation of PISA results, report writing and the production of tailored communication products to support the dissemination of PISA results and policy messages. These national reports and other communication products will present results in the context of the international PISA scales and include relevant analyses and information based on the policy priorities of each country. The reports will constitute a summary of key results and analysis designed to stimulate a constructive debate on improvement building upon and enriching already existing data and evidence from national, regional or international sources. The national reports will be the culmination of an engagement and communication strategy that is being implemented by each country, another new feature introduced by PISA-D. These strategies involve key stakeholders in each country in the survey and the discussion of the results and implications for policy. Stakeholders include pupils, parents, teachers, teacher unions, school principals, academia, civil society, media and central and local government.

Box 1.2 Key features of PISA-D

The content

The school-based survey assesses reading, mathematics and science, while the out-of-school survey includes reading and mathematics. PISA-D assesses not only whether students can reproduce knowledge, but also whether they can extrapolate from what they have learned and apply their knowledge in new situations. It emphasises the mastery of processes, the understanding of concepts, and the ability to function in various types of situations.

The students

Around 37 100 students will complete the school-based assessment, representing about 1 200 000 15-year-old students (in grades 7 or above) in the schools of the seven participating countries. Furthermore, around 16 200 students from five countries will participate in the out-of-school assessment, representing about 1 700 000 out-of-school youth between the ages of 14 and 16 and students aged 14 to 16 in grades 6 or below.

The assessment

The school-based assessment is a paper based test, lasting a total of two hours for each student. Test items are a mixture of multiple-choice questions and questions requiring students to construct their own responses. The items are organised in groups, each group based on a passage that sets out a real-life situation. The school-based assessment draws on about 195 test items, with different students taking different combinations of test items.

Students also answer a background questionnaire, which takes 35 minutes to complete. The questionnaire seeks information about the students themselves, their wellbeing, attainment, and engagement, their homes, their families, and their school and learning experiences. School principals complete a school questionnaire that describes the school, its students and teachers, and the learning environment. Teachers also complete a questionnaire about themselves, the school's resources, their teaching practice and their students.

The out-of-school assessment is conducted on a tablet computer. The test takes 50 minutes and test items are a mixture of multiple-choice questions and questions requiring respondents to construct their own responses. The items are organised in groups, each group based on a passage that sets out a real-life situation. Youth participating in the out-of-school assessment will answer about 38 test items, with different respondents taking different combinations of test items.

The out-of-school respondents also answer a background questionnaire, which takes about 30 minutes to complete. The questionnaire seeks information about the youth themselves, their wellbeing, attainment and attitudes towards learning, their homes and their school and learning experiences. Parents (or the most knowledgeable person) also answer a questionnaire about the youth's background and childhood experiences. A household observation questionnaire is completed by the interviewer and information about the location of the household is collected by PISA-D National Centres.

The PISA-D tests: School-based and out-of-school assessments

The PISA-D school-based instrument is a paper-based assessment designed as a two-hour test. This test design includes four clusters from each of the domains of science, reading and mathematics to measure trends. There are 12 different test booklets, each containing PISA 2015 trend items from two of the three core PISA domains. Each booklet allocated to students comprises four 30-minute clusters of test material. In total, students spend 120 minutes on each of the three subjects, reading, mathematics and science.

Each test booklet is completed by a sufficient number of students for appropriate estimates to be made of the achievement levels on all items by students in each country and in relevant subgroups within a country (such as boys and girls, and students from different social and economic contexts). Comparability with PISA 2015, a computer-based assessment, is assured through trend items. In addition, each student answers a 35-minute background questionnaire, which gathers contextual information that is analysed with the test results to provide a broader picture of student performance.

The PISA-D out-of-school instrument is a tablet-based assessment designed as a 50-minute test. The computer-based household survey Programme for the International Assessment of Adult Competencies (PIACC) was used as a model for choosing the delivery mode, and tablets were chosen over laptops on account of cost, efficiency and user friendliness. The test will include a 10-minute core module of basic reading and mathematics skills to ensure that respondents have an appropriate level of skills to proceed to the full assessment. An established minimum number of items answered correctly will determine the set of items that will be presented to respondents in the second stage of the cognitive assessment. The second stage was designed to take no longer than 40 minutes to complete. Respondents who pass the Core module will be randomly assigned to one of the 30 forms measuring reading and mathematical literacy. Respondents who fail the Core module will be directed to a 10-minute assessment of reading components followed by "Form 0", a 30-minute assessment of basic reading and mathematical literacy tasks. In addition, participants answer a 30-minute questionnaire.

Box 1.3 Paper-based or computer-based, does it make a difference?

There is a great deal of research on paper- and computer-based test performance, but findings are mixed. Some early studies indicated that reading speed was slower in a computer-based environment (Dillon, 1994) and less accurate (Muter et al., 1982), although these studies were conducted on proofreading tasks, not in an assessment situation. Richardson et al. (2002) reported that students found computer-based problem-solving tasks engaging and motivating, often despite the unfamiliarity of the problem types and the challenging nature of the items. They were sometimes distracted by attractive graphics, and sometime used poor heuristics when attempting tasks.

There is a large body of more recent literature on paper- and computer-based tests' equivalency (e.g. Macedo-Rouet et al., 2009; Paek, 2005); however these still reveal conflicting findings. In one of the largest comparisons of paper-based and computer-based testing, Sandene et al. (2008) found that eighth-grade students' mean score was four points higher on a computer-based mathematics test than on an equivalent paper-based test. Bennett et al. (2008) concluded from his research that computer familiarity affects performance on computer-based mathematics tests, while others have found that the range of functions available through computer-based tests can affect performance. For example, Mason, Patry and Berstein (2001) found that students' performance was negatively affected in computer-based tests compared to paper-based tests when there was no opportunity on the computer version to review and check responses. Bennett (2003) found that screen size affected scores on verbal-reasoning tests, possibly because smaller computer screens require scrolling.

By contrast, a meta-analysis of studies looking at K-12 students' mathematics and reading achievement (Wang et al., 2007) indicated that, overall, administration mode has no statistically significant effect on scores. A mode-effects study was conducted as part of the OECD Programme for the International Assessment of Adult Competencies (PIAAC) field trial. In this study, adults were randomly assigned to either a computer-based or paper-based assessment of literacy and numeracy skills. The majority of the items used in the paper delivery mode was adapted for computer delivery and used in this study. Analyses of these data reveal that almost all of the item parameters were stable across the two modes, thus showing that responses could be measured along the same literacy and numeracy scales (OECD, 2014). Given this evidence, it was hypothesised that PISA 2009 reading items could be transposed onto a screen for PISA 2015 without affecting trend data. This evidence was also the basis for hypothesising that PISA-D reading and mathematics items could be transposed onto a tablet without affecting trend data.

An overview of what is assessed in each domain

Box 1.4 presents definitions of the three domains assessed in PISA-D, which are the same definitions used for PISA 2015. The definitions all emphasise functional knowledge and skills that allow one to participate fully in society. Such participation requires more than just being able to carry out tasks imposed externally by, for example, an employer; it also means being able to participate in decision making. The more complex tasks in PISA-D require students to reflect on and evaluate material, not just to answer questions that have one correct answer.

Box 1.4 Definitions of the domains

Reading literacy: An individual's capacity to understand, use, reflect on and engage with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society.

Mathematical literacy: An individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens.

Scientific literacy: 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 which requires the competencies to explain phenomena scientifically, evaluate and design scientific enquiry, and interpret data and evidence scientifically.

Reading literacy (Chapter 2) is defined as an individual's ability to understand, use, reflect on and engage with written texts to achieve their goals, develop their knowledge and potential, and participate in society.

PISA-D assesses students' performance in reading through questions related to three major task characteristics:

- Processes, which refers to the cognitive approach that determines how readers engage with a text
- Text, which refers to the range of material that is read
- Situations, which refers to the range of broad contexts or purposes for which reading takes place.

Mathematical literacy (Chapter 3) is defined as individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens.

PISA-D assesses students' performance in mathematics through questions related to three interrelated aspects:

- The mathematical processes that describe what individuals do to connect the context of the problem with mathematics and thus solve the problem, and the capabilities that underlie those processes

- The mathematical content that is targeted for use in the assessment items

- The contexts in which the assessment items are located.

Scientific literacy (Chapter 4) is included in the school-based assessment only and is defined 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, which requires the competencies to explain phenomena scientifically, evaluate and design scientific enquiry, and interpret data and evidence scientifically.

PISA assesses students' performance in science through questions related to:

- Contexts: Personal, local/national and global issues, both current and historical, which demand some understanding of science and technology

- Knowledge: An understanding of the major facts, concepts and explanatory theories that form the basis of scientific knowledge. Such knowledge includes knowledge of both the natural world and 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 justification for their use (epistemic knowledge).

- Competencies: The ability to explain phenomena scientifically, evaluate and design scientific.

The evolution of reporting student performance in PISA and PISA-D

Results from PISA are reported using scales. Initially, the OECD average score for all three subjects was 500 with a standard deviation of 100, which meant that two-thirds of students across OECD countries scored between 400 and 600 points. These scores represent degrees of proficiency in a particular domain. In subsequent cycles of PISA, the OECD average score has fluctuated slightly around the original. The evolution of reporting student performance in PISA and PISA-D in the three domains is summarised in the sections below.

Reading literacy

Reading literacy was the major domain in 2000, and the reading scales were divided into five levels of knowledge and skills. The main advantage of this approach is that it is useful for describing what substantial numbers of students can do with tasks at different levels of difficulty. Results were also presented through three "aspect" subscales of reading: accessing and retrieving information; integrating and interpreting texts; and reflecting and evaluating texts. A proficiency scale was also available for mathematics and science, though without described levels.

PISA 2003 built upon this approach by specifying six proficiency levels for the mathematics scale. There were four "content" subscales in mathematics: space and shape, change and relationships, quantity, and uncertainty.

Similarly, the reporting of science in PISA 2006 specified six proficiency levels. The three "competency" subscales in science related to identifying scientific issues, explaining phenomena scientifically and using scientific evidence. Country performance was compared on the bases of knowledge about science and knowledge of science. The three main areas of knowledge of science were physical systems, living systems, and earth and space systems.

PISA 2009 marked the first time that reading literacy was re-assessed as a major domain. Trend results were reported for all three domains. PISA 2009 added a Level 6 to the reading scale to describe very high levels of reading proficiency. The bottom level of proficiency, Level 1, was relabelled as Level 1a. Another level, Level 1b, was introduced to describe the performance of students who would previously have been rated as "below Level 1", but who show proficiency in relation to new items that are easier than those included in previous PISA assessments. These changes allow countries to know more about what kinds of tasks students with very high and very low reading proficiency are capable of completing. To further extend the framework to the lower end of the scale of reading proficiency, PISA-D adds Level 1c to provide better coverage of basic processes, such as literal sentence and passage comprehension.

Mathematical literacy

Mathematics was re-assessed as a major domain in PISA 2012. In addition to the "content" subscales (with the "uncertainty" scale re-named as "uncertainty and data" for improved clarity), three new subscales were developed to assess the three processes in which students, as active problem solvers, engage. These three "process" subscales are: formulating situations mathematically; employing mathematical concepts, facts, procedures and reasoning; and interpreting, applying and evaluating mathematical outcomes (known as formulating, "employing" and "interpreting"). To further extend the framework to the lower end of the scale of mathematical literacy proficiency, PISA-D renames Level 1 as 1a, and creates two new proficiency levels at the lower end of the scale: Levels 1b and 1c to better measure basic processes, such as performing a simple calculation and selecting an appropriate strategy from a list.

Scientific literacy

Science, which was the main subject of assessment in PISA 2006, was again the main domain in PISA 2015. The assessment measures students' ability to: explain phenomena scientifically; evaluate and design scientific enquiry; and interpret data and evidence scientifically. The science scale was also extended by the addition of Level 1b to better describe the proficiency of students at the lowest level of ability who demonstrate minimal scientific literacy and who would previously not have been included in the reporting scales. To further extend the framework to the lower end of the scale of scientific literacy proficiency PISA-D adds Level 1c to gather information on basic skills at the lowest performance levels, such as being able to recall appropriate scientific knowledge but not apply such knowledge, or to make a simple prediction but not justify it.

The PISA-D context questionnaires

The focus of the PISA contextual questionnaires is on understanding how measures of student performance at age 15 are related to various aspects of school and classroom practice as well as other related factors, such as economic, social and cultural context. The PISA-D questionnaires include these aspects and also cover a broader set of well-being outcomes and a wider range of risk and protective factors, taking into account differences in life experiences of children in developing countries, both of those who are in school and of those who are not.

The contextual framework for PISA-D

The PISA-D questionnaire framework uses the Education Prosperity model (Willms, 2015) as an overarching framework, while also taking into account the goals of PISA-D, lessons from past PISA cycles and other international studies, recommendations from research literature and the priorities of the participating countries. Education prosperity, as applied in PISA-D, is a life-course approach that includes a core set of metrics for success at six key stages of development, covering the period from conception to adolescence. It identifies a key set of outcomes, called "Prosperity Outcomes" for six stages of development from conception to age 18, and a set of family, institutional and community factors, called "Foundations for Success", which drive these outcomes. PISA-D focuses on the fifth stage of the Educational Prosperity framework, late primary and lower secondary (ages 10 to 15).

The framework places great emphasis on equality and equity, with equality referring to differences among sub-populations in the distribution of their educational outcomes and equity referring to differences among sub-populations in their access to the resources and schooling processes that affect schooling outcomes. The PISA-D contextual framework also focuses on the measurement of socio economic status and poverty, with the purpose of exploring an international measure of poverty for youth in middle- and low- income countries while also extending the measure of the PISA index of economic, social and cultural status (ESCS).

The framework for the PISA-D questionnaires focuses on 14 modules of content. These modules measure the four Prosperity Outcomes, the five Foundations for Success, and the five demographic background factors relevant to assessing equality and equity that are listed below. In addition, the questionnaires include several teacher, school and system level background measures that provide a context for the Prosperity Outcomes. Chapter 5 presents the PISA-D questionnaire framework in detail.

Table 1.1 Modules assessed in the PISA-D questionnaires

1. Prosperity Outcomes	1.1 Academic performance (measured through the PISA-D tests)
	1.2 Educational attainment
	1.3 Health and well-being
	1.4 Student engagement
2. Foundation of Success	2.1 Inclusive environments
	2.2 Quality instruction
	2.3 Learning time
	2.4 Material resources
	2.5 Family and community support
3. Demographic factors for assessing equality and equity	3.1 Gender
	3.2 Disability
	3.3 Immigrant status
	3.4 Socioeconomic status and poverty
	3.5 Language spoken at home and language of instruction

PISA-D enhances the contextual questionnaires to better measure factors that are more strongly related to student performance in middle- and low-income countries while maintaining comparability with PISA on a set of core indicators. For example, the questionnaires collect more detailed data on students' language of instruction at school, language at home and their socioeconomic status, as measured by home possessions and parents' education, literacy skills and participation in the labour force. The questionnaires also identify additional indicators of educational success beyond performance on the PISA test. These indicators are measured, through questions about educational attainment, health and well-being, and student engagement in learning.

It is also important to note that the contextual information collected through the student, school and teacher questionnaires comprises only a part of the information available to PISA-D. System-level data describing the general structure of the education systems will be used in the PISA-D analysis and country reports. This system-level data includes information on the structure of national programmes, national assessments and examinations, instruction time, teacher training and salaries, educational finance (including enrolment), national accounts and population data. Available data on all of these indicators have been reviewed for PISA-D countries, identifying the current status of system-level data collection and availability in terms of quality and completeness (UNESCO Institute of Statistics, 2016).

The school-based questionnaires

The school-based questionnaires for students, teachers and the principals of schools have been developed in accordance with the contextual framework. These questionnaires take about 35 minutes for the students to complete and about 25 minutes for teachers and the principals. The responses to the questionnaires are analysed with the assessment results to provide at once a broader and more nuanced picture of student, school and system performance. These questionnaires seek information about:

- Students and their family backgrounds, including their economic, social and cultural capital, and the language they speak at home versus the language of instruction

- Aspects of students' lives, such as their level of educational attainment, their health and wellbeing, and their engagement with school

- Aspects of learning, including quality of instruction, inclusive environments, learning time, school material resources and family and community support

- Context of learning, including teacher, school and system level information

The out-of-school questionnaires

The out-of-school component questionnaires for youth, parents and interviewers have been developed in accordance with the contextual framework. These questionnaires each take between 15 and 30 minutes for the youth, the youth's most knowledgeable person (parent, guardian or other) and the interviewer respectively to complete. These questionnaires seek information about:

- Youth and their family backgrounds, including their economic, social and cultural capital, and the language they speak at home versus the language of instruction when they attended school

- Aspects of youths' lives, such as their level of educational attainment, their attitudes towards learning, their employment status, their habits and life outside of school, and their health and well-being

- Aspects of learning, including inclusive environments, family support, their perception of the inclusiveness of their school environment when they attended school, their reasons for being out of school and barriers preventing them from returning to school, and their family support and environment

- Aspects of youths’ early years, their educational experience and their parent/care-giver’s educational expectations for the youth
- Aspects of youths’ households, including location and surrounding characteristics

A collaborative project

PISA-D is a highly collaborative effort carried by the OECD Secretariat, contractors and nine participating countries with the support of several development partners and institutional partners.

The OECD’s Directorate for Education and Skills and the Development Co-operation Directorate share responsibility for the overall management of PISA-D, monitoring its implementation on a day-to-day basis and building consensus among countries. The OECD serves as the Secretariat and interlocutor between the PISA-D International Advisory Group (IAG), the PISA Governing Board (PGB), the Technical Advisory Group (TAG) and the PISA-D contractors. The OECD is also responsible for the capacity building of the participating countries, the production of the indicators, the analysis of results, and the preparation of the national reports and project publications in co-operation with the contractors and in close collaboration with the participating countries both at the policy level with the PGB and IAG and at the implementation level with the National Project Managers (NPMs).

The IAG, that is specific for PISA-D, meets annually and comprises government officials from participating countries, representatives of development partners supporting the initiative, representatives of institutional partners, such as UNESCO and UNICEF, invited experts and representatives of the OECD.

The PGB, representing all countries/economies with full PISA membership at senior policy levels, determines the policy priorities for PISA in the context of OECD objectives and oversees adherence to these priorities during the implementation of the programme. The PGB sets priorities for developing indicators, for establishing assessment instruments and for reporting results. Experts from participating countries/economies also serve on working groups to ensure that the instruments are internationally valid and take into account differences in the cultures and education systems.

The PISA-D TAG managed by the OECD explores technical issues that have policy or project implications and advises the OECD and its international contractors on these issues.

The PISA-D international contractors are responsible for survey operations and management and take the lead on supporting the countries to implement the programme. The contractors also take the lead on developing the enhanced assessment instruments, drawing on the technical expertise of the Subject Matter Expert Groups and Questionnaire Expert Groups that support PISA. The development of the PISA-D frameworks for reading, mathematics and science and the development of the PISA-D cognitive instruments are the responsibility of the contractor Educational Testing Service (ETS), while the design and development of the PISA-D questionnaires are the responsibility of the contractor The Learning Bar (TLB). Management and oversight of this survey, the development of the instruments, scaling and analysis are the responsibility of ETS as is the development of the electronic platform. Other partners or subcontractors involved with ETS include Pearson for the development of the cognitive frameworks, cApStAn for linguistic quality assurance and control and Westat for survey operations and sampling.

Participating countries implement the survey at the national level through National Centres (NCs). Within the NCs PISA is managed at the country level by National Project Managers (NPMs), subject to the agreed administration procedures and in accordance with the *PISA-D Technical Standards* put in place by the OECD and its contractors. The NPMs play a vital role in ensuring that implementation is of high quality and in accordance with the *PISA-D Technical Standards* help to shape and guide the project. They also verify and evaluate the survey results, analyses, reports and publications. The co-operation of students, teachers and principals in participating schools is crucial to the success of PISA-D during all stages of development and implementation. National experts from the participating countries contribute to the preparation of the frameworks and instruments and they also provide input for the design of analytical outputs. NCs collaborate with OECD on the analysis of PISA-D data for their countries and the production of national reports and other communication products.

From the outset of the project, the OECD has engaged the participation of the key international agencies and programmes concerned with student assessment and improving the quality of education in developing countries. These technical partners include UNESCO, UNESCO Institute of Statistics (UIS), the Education For All Global Monitoring Report (EFA GMR) team, UNICEF, the Global Partnership for Education (GPE) and the following assessment programmes: ASER; EGRA; EGMA; SACMEQ; PASEC; Pre-PIRLS and PIRLS; TIMSS; LLECE; STEP; LAMP; UWEZO; and WEI-SPS. Representatives of these agencies and programmes have been consulted on all aspects of project design and development.

The international and national costs of the project are being funded through a combination of development partner support and financing from the PISA-D countries. The development partners that have provided financing or aid-in-kind are France (AFD); Germany (BMZ/GIZ); GPE; Inter-American Development Bank (IADB); Ireland (Irish Aid); Japan (JICA); Korea; Microsoft Corporation; Positivo; Norway (Norad); Sunny Varkey Foundation; United Kingdom (DFID); and the World Bank.

Implementation of PISA-D

PISA-D is being implemented in five phases over the course of 2014 to 2019.

1. *Design, planning and co-ordination (2014-15)*: production of expert papers to inform the work of enhancing the assessment instruments, selection of international contractors to conduct the work, and the preparation of participating countries, including Capacity Needs Analysis, development of Capacity Building Plans and Project Implementation Plans. This phase also included the first and second annual meetings of the PISA-D International Advisory Group (IAG) and the first and second annual meetings of the PISA-D Technical Advisory Group (TAG) which were crucial for reaching agreements on the design of the initiative.

2. *Technical development (2015-16)*: review of assessment frameworks and items, selection of items, design of enhancements, preparation of materials, and planning for field trials, as well as the development of the project’s Analysis and Reporting Plan.

3. *Field trials and in-country data collection (2016-18)*: field trials in each country to test the enhanced instruments, reviewing and analysing the results of the field trial, preparation of materials for the main study data collection, and conducting the main study data collection.

4. *Analysis and report writing (2018-19)*: data cleaning and analysis, interpretation of results, eight countries writing their national reports supported by the OECD and its contractors.

5. *Report production, dissemination and post-pilot governance (2018-19)*: instruments finalised, an independent review of the project completed, national reports published, a project results report and a technical report published, a PISA-D international seminar, and PISA-D instruments incorporated in PISA from the 2021 cycle onwards.

Strengthening capacities

Nine countries (Bhutan, Cambodia, Ecuador, Guatemala, Honduras, Panama, Paraguay, Senegal and Zambia) have partnered with the OECD to develop and test the enhanced PISA instruments. With the exception of Panama, these countries have never before participated in PISA, but they have experience with regional or international assessments and conduct national student assessments.

In addition to delivering the enhancements to PISA discussed above, PISA-D also builds capacity for managing large-scale student learning assessment and using the results to support policy dialogue and decision making in the participating countries. The OECD offers participating countries training on a variety of topics, including framework and item development, sampling, translation/adaptation of survey instruments, data management, coding of students' responses, data analysis and reporting.

Participating countries have each established an NC and nominated an NPM to ensure appropriate infrastructure and resources are in place to implement the assessment in accordance with the PISA Technical Standards. A three-stage process has been developed and implemented to prepare countries for PISA-D participation:

1. Capacity Needs Analysis: ensures there is a solid foundational capacity for implementing the project and identifies areas of potential growth for the country.

2. Capacity Building Plan: addresses identified capacity needs and enhance the enabling environment for PISA, particularly the use of the results of assessments for national policy dialogue and evidence-based decision making.

3. Project Implementation Plan: describes the actions to be carried out by the specific entities and agents that are named and commissioned for implementation by the authorities of the participating country together with the necessary resources.

The project also promotes peer-to-peer learning by bringing together the countries already participating in PISA with PISA-D countries through individual country visits, staff exchanges, international meetings, technical trainings and workshops, and the development of country case studies. These country partnerships allow for sharing information about the implementation of the study and also about working with education stakeholders, using PISA to inform a broader national discussion about the value and standards of assessment, and preparing national reports and disseminating the assessment results.

PISA-D and the United Nation Sustainable Development Goals

The SDG Education 2030 agenda (UNESCO, 2015) that is set within the framework of the Sustainable Development Goals (United Nations, 2015) emphasises the quality, equity and measurement of learning outcomes for young children through to working adults. The challenge now is to define global learning indicators that can be measured and tracked on a global scale over time. Through its enhancement of PISA, the PISA-D initiative is designed to inform and support the monitoring, reporting and achievement of the Education SDG and its related targets and indicators, particularly those related to learning outcomes.

The OECD has been a key partner of UNESCO's and the other co-convening agencies in developing the Education SDG framework and works closely with UIS in the development of indicators that will be used to measure progress towards SDG achievement. In turn, UNESCO, UIS and the World Bank have partnered with the OECD in support of the PISA-D initiative.

One of the main challenges in implementing the SDG Education agenda is to define global learning indicators that can be measured and tracked on a global scale over time. The OECD, UIS and the World Bank are working together and with other key practitioners, policy makers, researchers, representatives of governments, civil society organisations, funders, UN agencies, and other stakeholders committed to improving learning outcomes in all countries – particularly, low- and middle-income countries. PISA-D and the OECD's plans for mainstreaming the outputs of the project in future cycles of PISA is a key contribution to these efforts and is an embodiment of international collaboration in support of the measurement and monitoring of learning outcomes in the context of the Education SDG.

REFERENCES

Adams, R. and Cresswell, J. (2016), "PISA for Development Technical Strand A: Enhancement of PISA Cognitive Instruments", *OECD Education Working Papers*, No.126, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jm5fb3f85s0-en.

Bennett, R.E. (2003), *Online Assessment and the Comparability of Score Meaning*, Research Memorandum, Educational Testing Service,Princeton, NJ.

Bennett, R.E. et al. (2008), "Does it matter if i take my mathematics test on computer? A second empirical study of mode effects in NAEP", *Journal of Technology, Learning, and Assessment*, Vol. 6/9.

Carr-Hill, R. (2015), "PISA for development technical strand c: Incorporating out-of-school 15-year-olds in the assessment", *OECD Education Working Papers*, No. 120, OECD Publishing, Paris, http://dx.doi.org/10.1787/5js0bsln9mg2-en.

Cresswell, J., U. Schwantner and C. Waters (2015), *A Review of International Large-Scale Assessments in Education: Assessing Component Skills and Collecting Contextual Data*, PISA, The World Bank, Washington, D.C./OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264248373-en.

Dillon, A. (1994), *Designing Usable Electronic Text: Ergonomic Aspects of Human Information Usage*, Taylor and Francis, London.

Lockheed, M., T. Prokic-Bruer and A. Shadrova (2015), *The Experience of Middle-Income Countries Participating in PISA 2000-2015*, PISA, The World Bank, Washington, D.C./OECD Publishing, Paris, 10.1787/9789264246195-en.

Macedo-Rouet, M. et al. (2009), "Students' performance and satisfaction with web vs. paper-based practice quizzes and lecture notes", *Computers and Education*, Vol. 53, pp. 375–384.

Mason, B., M. Patry and D. Berstein (2001), "An examination of the equivalence between non-adaptive computer based and traditional testing", *Journal of Education Computing Research*, Vol. 1/24, pp. 29-39.

Muter, P. et al. (1982), "Extended reading of continuous text on television screens", *Human Factors*, Vol. 24, pp. 501-508.

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

OECD (2014), *Technical Report of the Survey of Adult Skills (PIAAC)*, pre-publication, OECD, Paris, www.oecd.org/site/piaac/_Technical%20Report_17OCT13.pdf.

OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD Publishing, http://dx.doi.org/10.1787/9789264190511-en.

Paek, P. (2005), *Recent Trends in Comparability Studies: Pearson Educational Measurement*, Pearson Educational Measurement, http://images.pearsonassessments.com/images/tmrs/tmrs_rg/TrendsCompStudies.pdf (accessed 21 November 2007).

Richardson, M. et al. (2002), “Challenging minds? Students’ perceptions of computer-based world class tests of problem solving”, *Computers in Human Behavior*, Vol. 18/6, pp. 633-649.

Sandene, B. et al. (2005), *Online Assessment in Mathematics and Writing: Reports from the NAEP Technology-Based Assessment Project, Research and Development Series* (NCES 2005–457). US Department of Education, National Center for Education Statistics,US Government Printing Office, Washington, DC.

UNESCO Institute for Statistics (2016), *Making Education Count for Development: Data Collection and Availability in Six PISA for Development Countries*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264255449-en.

UNESCO (2015), Education 2030 Incheon Declaration and Framework for Action: Towards inclusive and equitable quality education and lifelong learning for all, available at: www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/ED/ED/pdf/FFA_Complet_Web-ENG.pdf.

United Nations (2015), Transforming our World – the 2030 Agenda for Sustainable Development, UN, New York, available at: www.un.org/pga/wp-content/uploads/sites/3/2015/08/120815_outcome-document-of-Summit-for-adoption-of-the-post-2015-development-agenda.pdf.

Wang, S. et al. (2007), “A meta-analysis of testing mode effects in Grade K–12 mathematics tests”, *Educational and Psychological Measurement*, Vol. 67, pp. 219-238.

Willms, J. D. (2015), *Educational Prosperity*, Fredericton: The Learning Bar Inc.

Willms, J. and Tramonte, L. (2015), “Towards the development of contextual questionnaires for the PISA for development study”, *OECD Education Working Papers*, No. 118, OECD Publishing, Paris, http://dx.doi.org/10.1787/5js1kv8crsjf-en.

CHAPTER 2. PISA FOR DEVELOPMENT READING FRAMEWORK

This chapter defines "reading literacy" as assessed in the Programme for International Student Assessment (PISA) and the extensions to the PISA reading framework that have been designed for the PISA for Development (PISA-D) project. It describes the processes involved in reading and the type of texts and response formats used in the PISA-D reading assessment and provides several sample items. The chapter also discusses how student performance in reading is measured and reported.

What is new in PISA-D? Extensions to the PISA reading literacy framework

The extensions made to the PISA frameworks for PISA-D are an attempt to gain more information about students at the bottom of the performance distribution, particularly for Level 1. The text in this chapter draws primarily on the PISA 2012 reading framework, with additions to address the extensions of the framework and some modifications to address aspects particularly important to assessment for PISA-D. Some specific elements from the 2018 framework have also been included.[1]The extensions occur primarily in four locations: the literature review, descriptions of the reading processes, descriptions of the proficiency levels, and discussion on assessing the proficiencies. The rationale behind these changes is also provided.

Reading literacy was the major domain assessed in 2000 for the first PISA cycle (PISA 2000). For the fourth PISA cycle (PISA 2009), it was the first to be revisited as a major domain, requiring a full review of its framework and new development of the instruments that represent it. For the seventh PISA cycle (2018), the conceptual framework for reading literacy is again being revised. This chapter discusses the conceptual framework underlying the PISA 2012 assessment of students' reading competencies and its extension to PISA-D. The definition of the domain is the same as in PISA 2009 (when it was assessed as the major domain for a second time), apart from the enhanced descriptions of the levels of competencies that fall below the current PISA Level 1.

Starting in 2009, the PISA reading literacy frameworks took into account digital reading, and the assessment of digital reading was implemented only as a computer-based assessment. Much of the content related to paper-based reading remained consistent across the 2009, 2012 and 2015 frameworks. However the 2015 framework was changed to make formulations for testing on computer. Because of this, the PISA-D framework is based on the 2012 framework. It must be stressed that both 2015 and 2018 offer a paper-based version that maintains its comparability with the computer-based version through the trend items. The use of trend items is the strategy used to ensure comparability between PISA-D and PISA 2015.

The PISA-D framework is designed for assessing the reading literacy of 15-year-old adolescents, which may be in or out of school. The 15-year-old students and youth need to read proficiently in order to participate in school activities (Shanahan and Shanahan, 2008). But most of them also use reading in a wide range of out-of-school contexts, for instance, to communicate with their peers, to acquire information related to their personal interests, or to interact with institutions and businesses (International Reading Association, 2012). Therefore, the framework must represent reading in a broad sense that encompasses basic as well as more advanced forms of reading, relevant for school as well as non-school situations. This includes not only the comprehension of a single given passage of text, but also an ability to find, select, interpret and evaluate information from the full range of texts associated with reading for school and out-of-school purposes.

The original reading literacy framework for PISA was developed through a consensus-building process involving reading experts selected by the participating countries to form the PISA 2000 reading expert group (REG). The definition of reading literacy evolved in part from the IEA Reading Literacy Study (1992) and the International Adult Literacy Survey (IALS, 1994, 1997, and 1998). In particular, it reflected the IALS emphasis on the importance of reading skills for active participation in society. It was also influenced by contemporary – and still current – theories of reading, which emphasise the multiple cognitive processes involved in reading and their interactive nature (Britt, Goldman and Rouet, 2012; Dechant, 1991; Rayner and Reichle, 2010; Rumelhart, 1985), models of discourse comprehension (Kintsch, 1998; Zwaan and Singer, 2003) and theories of performance in solving information problems (Kirsch, 2001; Kirsch and Mosenthal, 1990; Rouet, 2006).

Changes in our concept of reading since 2000 have already led to an expanded definition of reading literacy, which recognises motivational and behavioural characteristics of reading alongside cognitive characteristics. In light of recent research, reading engagement and metacognition were featured more prominently in the PISA 2009 reading literacy framework as elements that can make an important contribution to policy makers' understanding of factors that can be developed, shaped and fostered as components of reading literacy.

The PISA-D reading literacy framework provides additional emphasis on the basic components of the cognitive processes that underlie reading skills. These components include being able to locate information that is explicitly stated in text, to access and comprehend the meaning of individual words, and to understand the literal meaning of information as it is expressed in sentences as well as across passages. As such, these components can provide information about what these students *can do* with respect to the building blocks of reading literacy proficiency.

This chapter is organised into three major sections. The first section, "Defining reading literacy," explains the theoretical underpinnings of the PISA reading assessment, including the formal definition of the reading literacy construct. The second section, "Organising the domain of reading," describes three elements: processes, which refers to the cognitive approach that determines how readers engage with a text; text, which refers to the range of material that is read; and situation, which refers to the range of broad contexts or purposes for which reading takes place. The third section, "Assessing reading literacy", outlines the approach taken to apply the elements of the framework previously described, including factors affecting item difficulty, the response formats, coding and scoring, reporting proficiency, testing reading literacy among the out-of-school population and examples of items for addressing the extended PISA-D Framework.

Defining reading literacy

Definitions of reading and reading literacy have changed over time in parallel with changes in society, economy, and culture. The concept of learning, particularly the concept of lifelong learning, has expanded the perception of reading literacy. Literacy is no longer considered to be an ability acquired only in childhood during the early years of schooling. Instead, it is viewed as an expanding set of knowledge, skills and strategies that individuals build on throughout life in various contexts, through interaction with their peers and the wider community.

Cognitively based theories of reading emphasise the constructive nature of comprehension, the diversity of cognitive processes involved in reading and their interactive nature (Binkley, Rust and Williams 1997; Kintsch, 1998; McNamara and Magliano, 2009; Oakhill, Cain and Bryant, 2003; Rand Corporation, 2002; Zwaan and Singer, 2003). The reader generates meaning in response to text by using previous knowledge and a range of text and situational cues that are often socially and culturally derived. While constructing meaning, competent readers use various processes, skills and strategies to locate information, to monitor and maintain understanding (van den Broek, Risden and Husbye-Hartmann, 1995), and to critically assess the relevance and validity of the information (Richter and Rapp, 2014). These processes and strategies are expected to vary with context and purpose as readers interact with multiple continuous and non-continuous texts both in print and when using digital technologies (Britt and Rouet, 2012; Coiro, Knobel, Lankshear and Leu, 2008).

The PISA 2012 definition of reading literacy, the same used in PISA 2009 and 2015 and PISA-D, is shown in Box 2.1:

Box 2.1 The 2012 definition of reading literacy

Reading literacy is understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, develop one's knowledge and potential, and participate in society.

While this definition is taken for PISA-D also, the project extends the PISA definition of reading literacy through the incorporation of the concept of reading components. Reading components are the sub-skills, or building blocks, that underlie reading literacy (Oakhill, Cain and Bryant, 2003). As they develop and integrate, they facilitate proficient reading comprehension. Conversely, if the components are under-developed or deployed inefficiently, they may hinder a person's ability to comprehend texts (Perfetti, Landi and Oakhill, 2005). Although components can vary in their importance across languages (based on the structure of the language), there are several components that are generally agreed to be significant regardless of language family: word meaning (print vocabulary); sentence processing; and passage comprehension. An assessment of reading components was administered as part of the Programme for the International Assessment of Adult Competencies (PIAAC) and as an optional component of PISA 2012 (e.g. Sabatini and Bruce, 2009). The assessment of reading components can provide information on the component skills of students and out-of-school youth, particularly of those who fall at the lowest levels of literacy. They can also shed light on the kinds of educational/instructional programmes that improve their component skills, which will, in turn, improve their literacy. While word meaning is generally considered to be a proficiency that is already attained by 15 years, students in some countries may not have fully achieved this proficiency, most particularly when the established language of instruction is different from the student's home language. Thus, the PISA-D reading literacy framework incorporates the reading components of word comprehension, sentence processing and passage comprehension.

It should be noted that there are other critical reading components, including the visual recognition of the printed elements of the alphabet, decoding words into sounds and basic oral comprehension. These are not included as part of the PISA-D framework as they are assumed to be skills attained by 15-year-olds who attend school at their regular grade level as well as out-of-school 15-year-olds who have mastered these basic levels of literacy.

Box 2.2 Foundational reading skills required for PISA for Development

Successful performance on higher level reading tasks are dependent and built upon a foundation of component skills (e.g. Abadzi, 2003; Baer, Kutner and Sabatini, 2009; Curtis, 1980; Oakhill, Cain and Bryant, 2003; Perfetti, 2003; Rayner and Reichle, 2010; Sabatini and Bruce, 2009; Stine-Morrow, Miller and Hertzog, 2006). At the simplest view, reading consists of word recognition and linguistic comprehension, each being necessary but not sufficient for reading (e.g. Hoover and Tunmer, 1993). These components can be further elaborated to multiple foundational skills that are required to perform successfully at the lowest level of PISA. Below we outline five of these foundational skills. The first two are pre-conditions of the abilities needed to perform basic reading for PISA-D and are not assessed as part of the instrument. The remaining three are included as part of the assessment and are considered the basic skills that would be necessary to succeed at Level 1c.

1) Ability to relate characters (written symbols) to corresponding phonemes (acoustic sounds)

Reading requires mapping a system of printed visual symbols (individually and in combination) to the spoken form of the language (i.e. the phonetics, phonology) (e.g. Perfetti, 1985). However, there is a significant amount of variability in how the mapping is performed in different languages. For example, alphabetic writing systems map letters to phonemes, while other languages map characters at the syllable level and some at the level of individual words or morphemes (meaning-based units). Thus, the acquisition of this ability may vary by language.

2) Ability to recognise individual or groups of symbols as representing a word referring to objects and/or relationships

between words

The printed forms of objects and concepts given the particular orthographic and morphological structure of the language must also be recognized as representing meaningful words (e.g. Anderson and Freebody, 1981; Hirsch, 2003; McCutchen, Green and Abbott; 2008; Nagy and Townsend, 2012; Ouellet, 2006). It should be noted that this ability can differ across languages due to the orthographic differences across languages, the degree of regularity of the relationship between the print and oral language forms; and how morphological and grammatical/syntactical features of the language are encoded in words. For these reasons, it is difficult to ensure cross-language comparability in assessment, as this requires evaluating how to match the sources of difficulty in acquiring these print skills for each language, and balancing them across stimuli and tasks.

3) Ability to literally understand relationships among groups of words at the sentence level

An individual sentence serves as a complete unit of one or more coherent ideas (e.g. Kintsch, 1998), and a student must be able to comprehend the literal meaning of sentences of varying lengths. Reading a sentence requires both the syntactic processes that interpret the order and function of the words, and the semantic processes of interpreting the meaning of words and propositions (e.g. Kintsch, 1998; Rand Corporation, 2002).

4) Ability to literally understand explicit relationships between sentences at the level of short texts

Beyond individual sentences, a reader must be able to understand the literal meaning of passages of text. This requires forming a representation of the information contained across multiple sentences, connecting the idea units and structuring them in memory.

5) Ability to make low-level inferences about relationships across sentences of short texts

Students must be able to represent the information from connected sentences and infer specific relationships. These relationships can include connecting simple referents between one sentence and the next, such as the use of a nominal phrase in one sentence and a pronoun in the next, or creating coherence between two related sentences.

Reading literacy…

The term “reading literacy” is preferred to “reading” because it is likely to convey to a non-expert audience more precisely what the survey is measuring. “Reading” is often understood as simply decoding, or even reading aloud, whereas the intention of this survey is to measure something broader and deeper. Reading literacy includes a wide range of cognitive competencies, from basic decoding, to knowledge of words, grammar and larger linguistic and textual structures and features, to knowledge about the world.

In this assessment, “reading literacy” is intended to express the active, purposeful and functional application of reading in a range of situations and for various purposes. According to Holloway (1999), reading skills are essential to the academic achievement of middle and high school students. PISA assesses a wide range of students. Some will go on to university; some will pursue further studies in preparation for joining the labour force; some will enter the workforce directly after completing compulsory education. Achievement in reading literacy is not only a foundation for achievement in other subject areas within the education system, but also a prerequisite for successful participation in most areas of adult life (Cunningham and Stanovich, 1998; Smith, Mikulecky, Kibby and Dreher., 2000). Indeed, regardless of their academic or labour-force aspirations, students’ reading literacy is important for their active participation in their community and economic and personal life.

Reading literacy skills matter not just for individuals, but for economies as a whole. Policy makers and others are coming to recognise that in modern societies, human capital – the sum of what the individuals in an economy know and can do – may be the most important form of capital. Economists have for many years developed models showing generally that a country’s education levels are a predictor of its economic growth potential (Coulombe, Tremblay and Marchand, 2004).

…is understanding, using, reflecting on…

The word "understanding" is readily connected with the widely accepted concept of "reading comprehension", which emphasises that all reading involves some level of integrating information from the text with the reader's knowledge structures. In order to achieve some degree of understanding, the reader must decode written words, comprehend the literal meaning of sentences and passages, but also elaborate and reason about the information. Even the most basic forms of understanding require readers to draw on symbolic knowledge to identify words and make meaning from them. However, this process of integration can also be much broader, such as developing mental models of how texts relate to the world. The word "using" refers to the notions of application and function – doing something with what we read. "Reflecting on" is added to "understanding" and "using" to emphasise the notion that reading is interactive: readers draw on their own thoughts and experiences when engaging with a text. Of course, every act of reading requires some reflection, drawing on information from outside the text. Even at the earliest stages, readers draw on symbolic knowledge to decode a text and require some vocabulary knowledge to construct meaning. As readers develop their stores of information, experience and beliefs, they constantly, often unconsciously, test what they read against outside knowledge, thereby continually reviewing and revising their sense of the text.

…and engaging with…

A reading literate person not only has the skills and knowledge to read well, but also values and uses reading for a variety of purposes. It is therefore a goal of education to cultivate not only proficiency but also engagement in reading. Engagement in this context implies the motivation to read and comprises a cluster of affective and behavioural characteristics that include an interest in and enjoyment of reading, a sense of control over what one reads, involvement in the social dimension of reading, and diverse and frequent reading practices.

…written texts…

The term "written texts" is meant to include all those coherent texts in which language is used in its graphic form, whether printed and digital. Instead of the word "information", which is used in some other definitions of reading, the term "texts" was chosen because of its association with written language and because it more readily connotes literary as well as information-focused reading. The PISA-D reading literacy framework makes no assumption about the length or elaborateness of a written text. For example, a text could be a single word embedded within a graphic or short passages within a table.

These texts do not include aural language artefacts such as voice recordings; nor do they include film, TV, animated visuals or pictures without words. They do include visual displays such as diagrams, pictures, maps, tables, graphs and comic strips that include some written language (for example, captions). These visual texts can exist either independently or they can be embedded in larger texts. Digital texts are distinguished from printed texts in a number of respects, including physical readability; the amount of text visible to the reader at any one time; the way different parts of a text and different texts are connected with one another through hypertext links; and, given these text characteristics, the way that readers typically engage with digital texts. To a much greater extent than with printed or hand-written texts, readers need to construct their own pathways to complete any reading activity associated with a digital text.

…in order to achieve one's goals, develop one's knowledge and potential, and participate in society.

This phrase is meant to capture the full scope of situations in which reading literacy plays a role, from private to public, from school to work, from formal education to lifelong learning and active citizenship. "To achieve one's goals and to develop one's knowledge and potential" spells out the idea that reading literacy enables the fulfilment of individual aspirations – both defined ones, such as graduating or getting a job, and those less defined and less immediate that enrich and extend personal life and lifelong education. The word "participate" is used because it implies that reading literacy allows people to contribute to society as well as to meet their own needs. "Participating" includes social, cultural and political engagement.

Organising the domain of reading

This section describes how the domain is represented, a vital issue because the organisation and representation of the domain determines the test design and, ultimately, the evidence about student proficiencies that can be collected and reported.

Reading is a multidimensional domain. While many elements are part of the construct, not all can be taken into account in building the PISA assessment. Only those considered most important were selected.

The PISA reading literacy assessment is built on three major task characteristics to ensure a broad coverage of the domain:

- *processes*, which refers to the cognitive approach that determines how readers engage with a text
- *text*, which refers to the range of material that is read
- *situation*, which refers to the range of broad contexts or purposes for which reading takes place.

Note that the term *processes* – that is proposed as the term within PISA 2018 – is used in the PISA-D framework, although in PISA 2000 through PISA 2015, *processes* were referred to as *aspects*. This is because the term *processes* aligns better with the scholarly literature on reading comprehension and assessment. In addition the task characteristics are introduced in a different order than in the 2012 framework, in order to highlight those characteristics that are directly construct-relevant, as opposed to characteristics such as text types or task contexts, which are included mainly for purposes of coverage.

In PISA assessments, features of the text and processes variables (but not of the situation variable) are manipulated to influence the difficulty of a task. The processes are manipulated through the goals set in tasks.

Reading is a complex activity. The elements of reading do not exist independently of one another in neat compartments. The assignment of texts and tasks to framework categories does not imply that the categories are strictly partitioned or that the materials exist in atomised cells determined by a theoretical structure. The framework scheme is provided to ensure coverage, to guide the development of the assessment and to set parameters for reporting, based on what are considered the marked features of each task.

Processes

Processes are the mental strategies, approaches or purposes that readers use to negotiate their way into, around and between texts. Five processes were defined for PISA 2009-15 to guide the development of the reading literacy assessment tasks:

- retrieving information
- forming a broad understanding

- developing an interpretation
- reflecting on and evaluating the content of a text
- reflecting on and evaluating the form of a text.

For PISA-D, an additional process titled "literal comprehension" has been added. Literal comprehension requires students to comprehend explicitly stated information that may be found in individual words, sentences or passages. In addition, the concept of "retrieving information" is broadened to range from locating explicitly stated individual pieces of information, such as individual words or phrases, up to finding information in large passages.

As it is not possible to include sufficient items in PISA to report on each of the six processes as a separate subscale, for reporting on reading literacy these six processes are organised into three broad categories of processes:

- access and retrieve
- integrate and interpret
- reflect and evaluate.

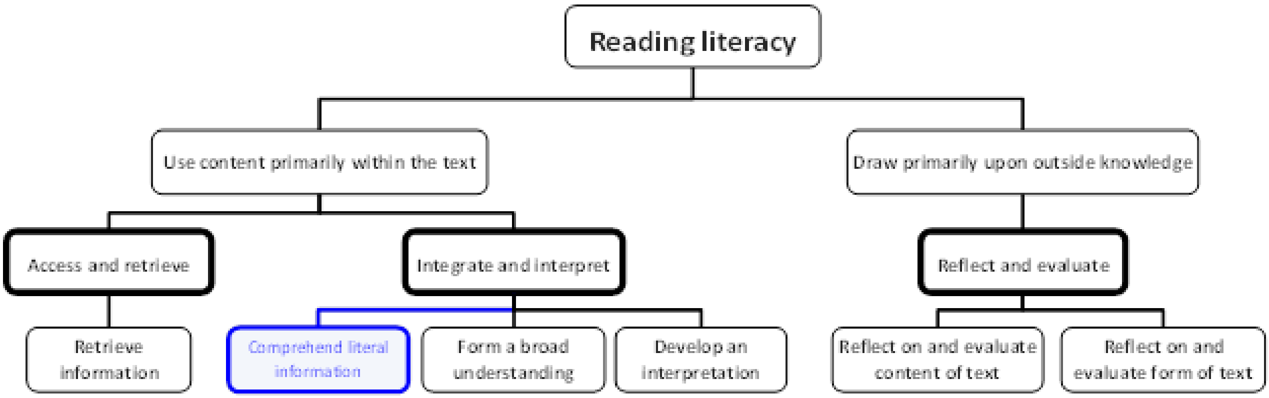
Generate literal comprehension, forming a broad understanding and *developing an interpretation* tasks focus the reader on relationships within a text. Tasks that focus on the whole text require readers to generate the literal meaning of words, individual sentences, and short passages. They also require forming a broad understanding; tasks that focus on relationships between parts of the text require developing an interpretation. The three are grouped together under *integrate and interpret*.

Tasks related with *retrieve information* process form the access and retrieve category.

Tasks addressing the last two processes, *reflecting on and evaluating the content of a text* and *reflecting on and evaluating the form of a text*, are grouped together into a single *reflect and evaluate* process category. Both require the reader to draw primarily on knowledge outside the text and relate it to what is being read. *Reflecting on and evaluating content* tasks are concerned with the notional substance of a text; *reflecting on and evaluating form* tasks are concerned with its structure or formal features.

Figure 2.1 shows the relationship between the five processes targeted in the test development for PISA in general and the additional process that will be assessed for PISA-D (in blue). The three broad categories reported on subscales in general PISA when reading literacy is the major domain are marked in bold. Because in PISA-D there is no major domain, reading literacy will be reported on a single overall scale only.

Figure 2.1 Processes targeted in reading literacy test development for PISA and PISA-D



An elaboration of the three broad process categories, encompassing tasks in both print and digital media, is given below.

Access and retrieve

Accessing and retrieving involves going to the information space provided and navigating in that space to locate and retrieve one or more distinct pieces of information. Access and retrieve tasks can range from locating individual pieces of information, such as the details required by an employer from a job advertisement, to finding a telephone number with several prefix codes, to finding a particular fact to support or disprove a claim someone has made.

While *retrieving* describes the process of selecting the required information, *accessing* describes the process of getting to the place, the information space, where the required information is located (e.g. see sample item 4, question 12.1). Both processes are involved in most *access and retrieve* tasks in PISA. However, some items may require retrieving information only, especially in the print medium where the information is immediately visible and where the reader only has to select what is appropriate in a clearly specified information space.

Difficulty will be determined by several factors, including the number of paragraphs or pages that need to be used, the amount of information to be processed on any given place, and the specificity and explicitness of the task directions.

Integrate and interpret

Integrating and interpreting involves processing what is read to construct an internal representation of the meaning of the text.

At the most basic levels of comprehension, readers need to be able to identify in print the meaning of individual words that would occur in the everyday listening lexicon of average adult speakers of the language (e.g. Sabatini and Bruce, 2009). This would include the everyday words of the language that would be used in common social and commerce situations, but not those specialised to technical or academic areas. Beyond the word level, students must be able to combine words in order to parse sentences and to represent their literal meaning. This involves an ability to acknowledge when sentences are ill-structured or simply do not make sense (see the section with sample items at the end of the chapter). Readers also need to combine the meaning of small sets of sentences in order to form internal representations of simple descriptions or narrations. Processing the literal meaning of a text is a foundational competency that then allows additional deeper processes to be performed on the text. In order to better represent this basic comprehension level, in PISA-D the category “integrate and interpret” is extended so as to include the process of “comprehending the literal meaning of text”. Tasks that specifically require this process will be included in the assessment.

Integrating focuses on demonstrating an understanding of the coherence of the *text* and involves the processes to make internal sense of a text. *Integrating* involves connecting various pieces of information to make meaning, whether it be identifying similarities and differences, making comparisons of degree, or understanding cause and effect relationships.

Interpreting also requires going beyond the literal meaning and refers to the process of making meaning from something that is not stated. When interpreting, a reader is identifying the underlying assumptions or implications of part or all of the text.

Both *integrating* and *interpreting* are required to *form a broad understanding.* A reader must consider the text as a whole or in a broad perspective. Students may demonstrate initial understanding by identifying the main topic or message or by identifying the general purpose or use of the text.

Both *integrating* and *interpreting* are also involved in *developing an interpretation*, which requires readers to extend their initial broad impressions so that they develop a deeper, more specific or more complete understanding of what they have read. *Integrating* tasks include identifying and listing supporting evidence, and comparing and contrasting information in which the requirement is to draw together two or more pieces of information from the text. In order to process either explicit or implicit information from one or more sources in such tasks, the reader must often infer an intended relationship or category. *Interpreting* tasks may involve drawing an inference from a local context: for example, interpreting the meaning of a word or phrase that gives a particular nuance to the text. This process of comprehension is also assessed in tasks that require the student to make inferences about the author’s intention, and to identify the evidence used to infer that intention.

The relationship between the processes of integration and interpretation may therefore be seen as intimate and interactive. Integrating involves first inferring a relationship within the text (a kind of interpretation), and then bringing pieces of information together, therefore allowing an interpretation to be made that forms a new integrated whole.

Reflect and evaluate

Reflecting and evaluating involves drawing upon knowledge, ideas or attitudes beyond the text in order to relate the information provided within the text to one’s own conceptual and experiential frames of reference.

Reflect items may be thought of as those that require readers to consult their own experience or knowledge to compare, contrast or hypothesise. *Evaluate* items are those that ask readers to make a judgment drawing on standards beyond the text.

Reflecting on and evaluating the content of a text requires the reader to connect information in a text to knowledge from outside sources. Readers must also assess the claims made in the text against their own knowledge of the world. Often readers are asked to articulate and defend their own points of view. To do so, readers must be able to develop an understanding of what is said and intended in a text. They must then test that mental representation against what they know and believe on the basis of either prior information, or information found in other texts. Readers must call on supporting evidence from within the text and contrast it with other sources of information, using both general and specific knowledge as well as the ability to reason abstractly.

Reflecting on and evaluating the form of a text requires readers to stand apart from the text, to consider it objectively, and to evaluate its quality and appropriateness. Implicit knowledge of text structure, the style typical of different kinds of texts, can play an important role in these tasks. Evaluating how successful an author is in portraying some characteristic or persuading a reader depends not only on substantive knowledge but also on the ability to detect subtleties in language.

To some extent every critical judgment requires the reader to consult his or her own experience; some kinds of reflection, on the other hand, do not require evaluation (for example, comparing personal experience with something described in a text). Thus evaluation might be seen as a subset of reflection.

The processes of reading in print and digital-media

The three broad processes defined for PISA reading literacy are not conceived of as entirely separate and independent, but rather as interrelated and interdependent. Indeed from a cognitive processing perspective, they can be considered semi-hierarchical: it is not possible to interpret or integrate information without having first retrieved it. And it is not possible to reflect on or evaluate information without having made some sort of interpretation. In PISA, however, the framework description of reading processes distinguishes approaches to reading that are demanded for different contexts and purposes; these are then reflected in assessment tasks that emphasise one or other process.

For PISA-D, the distribution of tasks across the major framework variables of situation and text should closely mirror the distributions used for the print items in PISA 2012, both for the school-based and the out-of-school tests. The distribution of process variables does have some differences.

Table 2.1 shows the approximate distribution of reading score points by the processes for the PISA 2012 assessment and the desired distribution of reading score points by the processes for PISA-D. Note that the distribution puts greater emphasis on *access and retrieve*, most particularly at the lower levels of proficiency, while also putting lower emphasis on *reflect and evaluate*. This enhances the sensitivity to competencies that will tend to fall at the lower levels of the PISA scale.

Table 2.1 Distribution of score points in reading, by processes, for PISA 2012 (approximate distribution) and PISA-D (desired distribution)

Processes (aspects)	Percentage of total score points in PISA 2012: print	Percentage of total score points in PISA-D
Access and retrieve	22	25-30% with 15% below Level 3
Integrate and interpret	56	45-55%
Reflect and evaluate	22	15-25%
Complex	0	0
Total	**100**	**100**

The desired distribution specifies the blueprint for the selection of items according to important aspects of the domain frameworks. Item selection is based on the assessment design as well as item characteristics related to a number of framework aspects, including coding requirement, process, situation, and text format, as well as consideration of the items' psychometric properties and appropriateness for this assessment. Following the assessment, the actual distributions of items across the framework aspects will be described in relation to the desired distributions. The extent to which the item pool for the assessment meets the framework specifications will be discussed in the technical report in the context of practical constraints in the item selection process.

Situation

The PISA-D situation variables remain the same as those for PISA 2012. They were adapted from the Common European Framework of Reference (CEFR) developed for the Council of Europe (Council of Europe, 1996). The four situation variables – personal, public, educational and occupational – are described in the following paragraphs.

The *personal* situation relates to texts that are intended to satisfy an individual's personal interests, both practical and intellectual. This category also includes texts that are intended to maintain or develop personal connections with other people. It includes personal letters, fiction, biography and informational texts that are intended to be read to satisfy curiosity, as a part of leisure or recreational activities. In the digital medium it includes personal emails, instant messages and diary-style blogs.

The *public* category describes the reading of texts that relate to activities and concerns of the larger society. The category includes official documents and information about public events. In general, the texts associated with this category assume a more or less anonymous contact with others; they also therefore include forum-style blogs, news websites, and public notices that are encountered both on line and in print.

The content of *educational* texts is usually designed specifically for the purpose of instruction. Printed text books and interactive learning software are typical examples of material generated for this kind of reading. Educational reading normally involves acquiring information as part of a larger learning task. The materials are often not chosen by the reader, but instead assigned by an instructor. The model tasks are those usually identified as "reading to learn" (Sticht, 1975; Stiggins, 1982).

Many 15-year-olds will move from school into the labour force within one to two years, and many out-of-school youth may already be part of the work force. A typical *occupational* reading task is one that involves the accomplishment of some immediate task. It might include searching for a job, either in a print newspaper's classified advertisement section, or on line; or following workplace directions. The model tasks of this type are often referred to as "reading to do" (Sticht, 1975; Stiggins, 1982).

Situation is used in PISA reading literacy to define texts and their associated tasks, and refers to the contexts and uses for which the author constructed the text. The manner in which the situation variable is specified is therefore about supposed audience and purpose, and is not based simply on the place where the reading activity is carried out. Many texts used in classrooms are not specifically designed for classroom use. For example, a piece of literary text may typically be read by a 15-year-old in a mother-tongue language or literature class, yet the text was written (presumably) for readers' personal enjoyment and appreciation. Given its original purpose, such a text is classified as *personal* in PISA. As Hubbard (1989) has shown, some kinds of reading usually associated with out-of-school settings for children, such as rules for clubs and records of games, often take place unofficially at school as well. These texts are classified as *public* in PISA. Conversely, textbooks are read both in schools and in homes, and the process and purpose probably differ little from one setting to another. Such texts are classified as *educational* in PISA.

It should be noted that the four categories overlap. In practice, for example, a text may be intended both to delight and to instruct (personal and educational); or to provide professional advice that is also general information (occupational and public). While content is not a variable that is specifically manipulated in this study, by sampling texts across a variety of situations the intent is to maximise the diversity of content that will be included in the PISA reading literacy survey.

Table 2.2 shows the approximate distribution of score points by situation for print reading tasks in PISA 2012 and the desired distribution for PISA-D. The distributions of situations used in PISA 2012 can be maintained at the same approximate values for PISA-D.

Table 2.2 Distribution of score points in reading, by situation, for PISA 2012 (approximate distribution) and PISA-D (desired distribution)

Situation	**Percentage of total score points in PISA 2012: print**	**Percentage of total score points in PISA-D**
Personal	36	25-45
Educational	33	25-45
Occupational	20	15-25
Public	11	5-15
Total	**100**	**100**

Text

The text dimensions for PISA-D remain the same as those used in PISA 2012. Reading requires material for the reader to read. In an assessment, that material – a text (or a set of texts) related to a particular task – must be coherent within itself. That is, the text must be able to stand alone without requiring additional material to make sense to the proficient reader. While it is obvious that there are many different kinds of texts and that any assessment should include a broad range, it is not so obvious that there is an ideal categorisation of kinds of texts. The addition of digital reading to the framework has made this issue still more complex. In 2009 and 2012, there have been four main text classifications[2]:

- Medium: print and digital
- Environment: authored and message-based
- Text format: continuous, non-continuous, mixed and multiple
- Text type: description, narration, exposition, argumentation, instruction and transaction

The classification of medium – print and digital – is applied to each text as the broadest distinction. Below that classification, the text format and text type categories are applied to all texts, whether print or digital. The environment classification, on the other hand, is only applicable to digital texts.

Medium

Since PISA 2009, an important major categorisation of texts is the classification by medium: print or digital.

Print text usually appears on paper in forms such as single sheets, brochures, magazines and books. The physical status of the printed text encourages (though it does not compel) the reader to approach the content of the text in a particular sequence. In essence, printed texts have a fixed or static existence. Moreover, in real life and in the assessment context, the extent or amount of the text is immediately visible to the reader.

Digital text may be defined as the display of text through liquid crystal display (LCD), plasma, thin film transistor (TFT), and other electronic devices. For the purposes of PISA, however, digital text is synonymous with *hypertext*: a text or texts with navigation tools and features that make possible and indeed even require non-sequential reading. Each reader constructs a “customised” text from the information encountered at the links he or she follows. In essence, such digital texts have an unfixed, dynamic existence. In the digital medium, typically only a fraction of the available text can be seen at any one time, and often the extent of text available is unknown. The PISA-D instruments do not include hypertext, but digital text is mentioned here for completeness.

Text format

An important classification of texts is the distinction between continuous and non-continuous texts.

Texts in *continuous* and *non-continuous* format appear in both the print and digital media. *Mixed* and *multiple* format texts are also prevalent in both media, particularly so in the digital medium. Each of these four formats is elaborated as follow:

Continuous texts are formed by sentences organised into paragraphs. These may fit into even larger structures, such as sections, chapters, and books (e.g. newspaper reports, essays, novels, short stories, reviews and letters for the print medium, and reviews, blogs and reports in prose for the digital).

Non-continuous texts are organised differently to *continuous* texts, and therefore require a different kind of reading approach. *Non-continuous* texts are most frequently organised in matrix format, composed of a number of lists (Kirsch and Mosenthal, 1990) (e.g. lists, tables, graphs, diagrams, advertisements, schedules, catalogues, indexes and forms).

Many texts in both print and digital media are single, coherent artefacts consisting of a set of elements in both a *continuous* and *non-continuous* format. In well-constructed *mixed* texts, the constituents (e.g. a prose explanation, along with a graph or table) are mutually supportive through coherence and cohesion links at the local and global level. *Mixed* text in the print medium is a common format in magazines, reference books and reports. In the digital medium, authored web pages are typically mixed texts, with combinations of lists, paragraphs of prose, and often graphics. Message-based texts such as online forms, e-mail messages and forums also combine texts that are *continuous* and *non-continuous* in format.

Multiple texts are defined as those that have been generated independently, and make sense independently; they are juxtaposed for a particular occasion or may be loosely linked together for the purposes of the assessment. The relationship between the texts may not be obvious; they may be complementary or may contradict one another. For example, with digital texts, a set of websites from different companies providing travel advice may or may not provide similar directions to tourists. For paper-based texts, multiple texts may include a bus time schedule, a map and a text explaining a set of tours around a town. Multiple texts may have a single "pure” format (for example, continuous), or may include both continuous and non-continuous texts.

Table 2.3 shows the approximate distributions of score points for print reading tasks by text format for PISA 2012, which should be maintained for PISA-D.

Table 2.3 Distribution of score points in reading, by text format, for PISA 2012 (approximate distribution) and PISA-D (desired distribution)

Text format	Percentage of total score points in PISA 2012: print	Percentage of total score points in PISA-D
Continuous	58	50-60
Non-continuous	31	25-35
Mixed	9	5-15
Multiple	2	0-10
Total	**100**	**100**

Text type

A different categorisation of text is by text type: description, narration, exposition, argumentation, instruction, and transaction. The text types are maintained the same for PISA-D as they have been since PISA 2009.

Texts as they are found in the world typically resist categorisation; they are usually not written with rules in mind, and tend to cut across categories. That notwithstanding, in order to ensure that the reading instrument samples across a range of texts that represent different types of reading PISA categorises texts based on their predominant characteristics,.

The following classification of texts used in PISA is adapted from the work of Werlich (1976).

Description is the type of text where the information refers to properties of objects in space. The typical questions that descriptive texts provide an answer to are *what* questions (e.g. a depiction of a particular place in a travelogue or diary, a catalogue, a geographical map, an online flight schedule, or a description of a feature, function or process in a technical manual).

Narration is the type of text where the information refers to properties of objects in time. Narration typically answers questions relating to *when*, or *in what sequence*. Why characters in stories behave as they do is another important question that narration typically answers (e.g. a novel, a short story, a play, a biography, a comic strip, fictional texts and a newspaper report of an event). The proportion of narrative texts in the print medium in PISA 2012 was a little greater than that in the previous PISA cycles (2000-09), at about 20% (formerly about 15%).

Exposition is the type of text in which the information is presented as composite concepts or mental constructs, or those elements into which concepts or mental constructs can be analysed. The text provides an explanation of how the different elements interrelate in a meaningful whole and often answers questions about *how* (e.g. a scholarly essay, a diagram showing a model of memory, a graph of population trends, a concept map and an entry in an online encyclopaedia).

Argumentation is the type of text that presents the relationship among concepts or propositions. Argument texts often answer *why* questions. An important subclassification of argument texts is persuasive and opinionative texts, referring to opinions and points of view. Examples of text in the text type category *argumentation* are a letter to the editor, a poster advertisement, the posts in an online forum, and a Web-based review of a book or film.

Instruction is the type of text that provides directions on what to do. The text presents directions for certain behaviours in order to complete a task (e.g. a recipe, a series of diagrams showing a procedure for giving first aid and guidelines for operating digital software).

Transaction represents the kind of text that aims to achieve a specific purpose outlined in the text, such as requesting that something is done, organising a meeting or making a social engagement with a friend. Before the spread of digital communication, this kind of text was a significant component of some kinds of letters and, as an oral exchange, the principal purpose of many phone calls. This text type was not included in Werlich's (1976) categorisation. It was used for the first time in the PISA 2009 framework because of its prevalence in the digital medium (e.g. everyday email and text message exchanges between colleagues or friends that request and confirm arrangements).

Strategy to extend the framework to provide better coverage of basic literacy levels

Two strategies are used in order to extend the framework to lower levels of reading proficiencies. Firstly, additional item types are included to assess word meaning, basic sentence and passage comprehension, and literal meaning (see above the description of the process "integrate and interpret"). The purpose of the tasks is to measure the extent to which students understand the literal and inferential meaning of words and connected text. Two tasks were defined: *sentence processing* and *passage comprehension*.

The sentence processing tasks assess the ability to comprehend written sentences of varying lengths. In the PISA reading components assessment, the construct is instantiated in a sensibility judgment task. Its purpose is to measure the extent to which students can comprehend sentences of increasing lengths. In the task, students see a set of sentences and decide if they make sense (YES) or do not make sense (NO) with respect to general knowledge about the real world, as in the first item in the section on sample items, or the internal logic of the sentence itself, as in the second item of sample item 1.

The basic passage comprehension tasks assess the ability to understand the literal, "gist" meaning of connected text and to make low-level inferences across sentences in the text. In the PISA and PIAAC reading components assessments, the construct has been instantiated in an embedded cloze task. Its purpose is to measure the extent to which students understand the literal and inferential meaning of connected text. In the task, the participant sees a passage in which the sentences include an embedded cloze item (two word choices are given for a single blank). The participant reads the passage silently and circles the word that correctly completes each sentence. Sample item 2 shows an example of a passage comprehension task with multiple items embedded within it.

The second strategy consists in adapting existing PISA tasks to assess low-level comprehension and access and retrieve processes. Sample items 3, 4 and 5 illustrate this.

Assessing reading literacy

The previous section outlined the conceptual framework for reading literacy. The concepts in the framework must in turn be represented in tasks and questions in order to collect evidence of students' proficiency in reading literacy.

The distribution of tasks across the major framework variables of process, situation and text was discussed in the previous section. In this section, the framework describes the distribution of tasks across the major framework variables as well as some of the other major issues in constructing and operationalising the assessment: factors affecting item difficulty and how difficulty can be manipulated; the choice of response formats; issues around coding and scoring; strategy to extend the framework to provide better coverage of basic literacy levels; reporting proficiency in reading in PISA-D; testing reading literacy among the out-of-school population; and examples of items for addressing the extended PISA-D framework.

Factors affecting item difficulty

The difficulty of any reading literacy task depends on an interaction among several variables. Drawing on Kirsch and Mosenthal's work (e.g. Kirsch, 2001; Kirsch and Mosenthal, 1990), we can manipulate the difficulty of items by applying knowledge of the following process and text format variables.

In *access and retrieve* tasks, difficulty is conditioned by the number of pieces of information that the reader needs to locate, by the amount of inference required, by the amount and prominence of competing information, and by the length (e.g. number of words, sentences, paragraphs) and complexity of the text.

In *integrate and interpret* tasks, difficulty is affected by the type of interpretation required (for example, making a comparison is easier than finding a contrast, and comprehending a specified causal link is easier than inferring an implicit causal relationship); by the number of pieces of information to be considered; by the degree and prominence of competing information in the text; and by the nature of the text. The less familiar and the more abstract the content and the longer and more complex the text, the more difficult the task is likely to be.

In *reflect and evaluate* tasks, difficulty is affected by the type of reflection or evaluation required (from least to most difficult, the types of reflection are: connecting; explaining and comparing; hypothesising and evaluating); by the nature of the knowledge that the reader needs to bring to the text (a task is more difficult if the reader needs to draw on narrow, specialised knowledge rather than broad and common knowledge); by the relative abstraction and length of the text; and by the depth of understanding of the text required to complete the task.

In tasks relating to *continuous texts*, difficulty is influenced by the length of the text, the explicitness and transparency of its structure, how clearly the parts are related to the general theme, and whether there are text features, such as paragraphs or headings, and discourse markers, such as sequencing words.

In tasks relating to *non-continuous texts*, difficulty is influenced by the amount of information in the text; the list structure (simple lists are easier to negotiate than more complex lists); whether the components are ordered and explicitly organised, for example with labels or special formatting; and whether the information required is in the body of the text or in a separate part, such as a footnote.

Response formats

Coding requirements are shown in Table 5 for print score points in relation to the three processes of reading literacy and for digital score points in relation to the four processes. Items that require expert judgment consist of open-constructed and short-constructed responses that require expert coding. Items that do not require coder judgment consist of multiple-choice, complex multiple-choice and closed-constructed response items. The closed-constructed response items are those that require the student to generate a response, but require minimal judgment on the part of a coder.

Distribution of coding requirements for PISA-D should be kept comparable to mainstream PISA assessments. The distribution of item types in print reading does not vary much from one cycle/administration to the next. However, the selection for 2012 has a slightly higher proportion of items that do not require expert coding than in previous cycles: 58% non-expert coded and 42% expert coded in 2012 (compared with 55% and 45% respectively in previous administrations). The same ratio applies to print and to digital reading in PISA 2012.

Table 2.4 shows the approximate distribution of score points by coding requirement for each reading process in PISA 2012 and in the paper-based PISA-D test. Due to the extra testing time it would involve, the tablet based test does not include items that require expert judgement in coding.

Table 2.4 Distribution of score points in reading, by coding requirement for each reading process, in PISA 2012 (approximate distribution) and PISA-D (desired distribution)

	Percentage of total score points in PISA 2012: Print reading			Percentage of total score points in PISA-D		
Process (aspect)	**Expert judgement required**	**No expert judgement required**	**Total**	**Expert judgement required**	**No expert judgement required**	**Total**
Access and retrieve	4	18	22	0-10	10-20	20-30
Integrate and interpret	20	36	56	15-30	30-40	45-60
Reflect and evaluate	18	4	22	15-25	0-10	20-30
Complex	0	0	0	0	0	0
TOTAL	**42**	**58**	**100**	**35-50**	**45-65**	**100**

Coding and scoring

Codes are applied to test items, either by a more or less automated process of capturing the alternative chosen by the student for a multiple-choice answer, or by a human judge (expert coder) selecting a code that best captures the kind of response given by a student to an item that requires a constructed response. The code is then converted to a score for the item. For multiple-choice or closed-response format items, the student has either chosen the designated correct answer or not, so the item is scored as 1 (full credit) or 0 (no credit) respectively. For more complex scoring of constructed response items, some answers, even though incomplete, indicate a higher level of reading literacy than inaccurate or incorrect answers, and receive partial credit.

Reporting proficiency in reading in PISA-D

PISA reports results in terms of proficiency scales that are interpretable for the purposes of policy. To capture the progression of complexity and difficulty, from 2009 and up to 2018, PISA has used seven levels based on the PISA 2009 combined print reading literacy scale. For PISA-D, an additional level has been added at the lowest level, so the combined print reading literacy scale is divided into eight levels. Figure 2 describes these eight levels of print reading proficiency. Level 6 is the highest described level of proficiency (Level 5 was the highest level before PISA 2009 reading assessments). Levels 2, 3, 4 and 5 remain the same as in PISA 2000. In the mainstream PISA, the lowest bottom level of measured proficiency is Level 1b, with Level 1a being the second lowest level. For PISA-D, Level 1c is added as the lowest level of proficiency with a focus on understanding words, short phrases and extracting literal meaning from sentences. These different levels of proficiency allow countries to know more about the kinds of tasks students with very high and very low reading proficiency are capable of performing.

Figure 2.2 Summary description of the eight levels of reading proficiency in PISA-D

Level	Lower score limit	Percentage of students across OECD countries at each proficiency level in PISA 2012	Percentage of students across 18 middle- and low-income countries at each proficiency level in PISA 2012	Characteristics of tasks
6	698	1.1%	0.1%	Tasks at this level typically require the reader to make multiple inferences, comparisons and contrasts that are both detailed and precise. They require demonstration of a full and detailed understanding of one or more texts and may involve integrating information from more than one text. Tasks may require the reader to deal with unfamiliar ideas, in the presence of prominent competing information, and to generate abstract categories for interpretations. *Reflect and evaluate* tasks may require the reader to hypothesise about or critically evaluate a complex text on an unfamiliar topic, taking into account multiple criteria or perspectives, and applying sophisticated understandings from beyond the text. A salient condition for *access and retrieve* tasks at this level is precision of analysis and fine attention to detail that is inconspicuous in the texts.
5	626	7.3%	1.1%	Tasks at this level that involve retrieving information require the reader to locate and organise several pieces of deeply embedded information, inferring which information in the text is relevant. Reflective tasks require critical evaluation or hypothesis, drawing on specialised knowledge. Both interpretative and reflective tasks require a full and detailed understanding of a text whose content or form is unfamiliar. For all processes of reading, tasks at this level typically involve dealing with concepts that are contrary to expectations.
4	553	21.0%	6.7%	Tasks at this level that involve retrieving information require the reader to locate and organise several pieces of embedded information. Some tasks at this level require interpreting the meaning of nuances of language in a section of text by taking into account the text as a whole. Other interpretative tasks require understanding and applying categories in an unfamiliar context. Reflective tasks at this level require readers to use formal or public knowledge to hypothesise about or critically evaluate a text. Readers must demonstrate an accurate understanding of long or complex texts whose content or form may be unfamiliar.
3	480	29.1%	19.1%	Tasks at this level require the reader to locate, and in some cases recognise the relationship between, several pieces of information that must meet multiple conditions. Interpretative tasks at this level require the reader to integrate several parts of a text in order to identify a main idea, understand a relationship or construe the meaning of a word or phrase. They need to take into account many features in comparing, contrasting or categorising. Often the required information is not prominent or there is much competing information; or there are other text obstacles, such as ideas that are contrary to expectation or negatively worded. Reflective tasks at this level may require connections, comparisons, and explanations, or they may require the reader to evaluate a feature of the text. Some reflective tasks require readers to demonstrate a fine understanding of the text in relation to familiar, everyday knowledge. Other tasks do not require detailed text comprehension but require the reader to draw on less common knowledge.
2	407	23.5%	30.3%	Some tasks at this level require the reader to locate one or more pieces of information, which may need to be inferred and may need to meet several conditions. Others require recognising the main idea in a text, understanding relationships, or construing meaning within a limited part of the text when the information is not prominent and the reader must make low level inferences. Tasks at this level may involve comparisons or contrasts based on a single feature in the text. Typical reflective tasks at this level require readers to make a comparison or several connections between the text and outside knowledge, by drawing on personal experience and attitudes.
1a	335	12.3%	25.9%	Tasks at this level require the reader to locate one or more independent pieces of explicitly stated information; to recognise the main theme or author's purpose in a text about a familiar topic, or to make a simple connection between information in the text and common, everyday knowledge. Typically the required information in the text is prominent and there is little, if any, competing information. The reader is explicitly directed to consider relevant factors in the task and in the text.

1b	262	4.4%	12.6%	Tasks at this level require the reader to locate a single piece of explicitly stated information in a prominent position in a short, syntactically simple text with a familiar context and text type, such as a narrative or a simple list. The text typically provides support to the reader, such as repetition of information, pictures or familiar symbols. There is minimal competing information. In tasks requiring interpretation the reader may need to make simple connections between adjacent pieces of information.
1c	*	1.3% (percentage of students scoring below Level 1b in PISA 2012)	4.3% (percentage of students scoring below Level 1b in PISA 2012)	Tasks at this level require a reader to understand the meaning of individual written words and short phrases. The tasks require students to locate a single word or phrase in a short list or text, to recognise the printed forms of common objects and concepts, or to extract the literal meaning of individual sentences and very short syntactically simple passages with familiar contexts. Texts support students with explicit pointers to the information and with repetition, pictures or familiar symbols with limited competing information.

Notes: * Will be available after the main study.
Descriptors 1b through 6 are the same as those used in PISA 2012 and 2015.

Testing reading literacy among the out-of-school population

The extended Reading PISA-D framework is appropriate for 15-year-old students whether in or out of school. The units and items are not directly school contextually-based, and thus, there is no particular requirement or change needed in the units that are categorized as relevant for educational activities, since educational activities also occur out-of-school. Therefore, the distribution and selection of units and items can be maintained the same for PISA-D in-school and out-of-school populations.

The out-of-school component is assessed on a tablet computer, but only fixed-text items are used, so it is appropriate to use the same framework as for the paper based test.

Box 2.3 Delivery mode

The PISA-D school based assessment is paper based, while the out-of-school assessment is conducted on a tablet computer. To ensure comparability between the tests, the tablet based instruments for PISA-D are formed by a subgroup of the items used for the paper based assessment. All of these items were originally designed for a paper based assessment, so when moving to a tablet based delivery, care was taken to maintain comparability between the assessments. The PISA 2015 framework describes some factors that must be considered when transposing items from paper to computer mode. These elements were also taken into account when designing the out-of-school instrument for PISA-D.

Item types: The computer provides a range of opportunities for designers of test items, including new item formats (e.g. drag-and-drop, hotspots). Since the PISA-D tablet bases tests uses a sub group of items form the paper based test, there is less opportunity to exploit innovative item types and the majority of response formats remains unchanged.

Stimulus presentation: A feature of fixed texts defined in the construct is that “the extent or amount of the text is immediately visible to the reader”. Clearly, it is impossible, both on paper and on a screen, to have long texts displayed on a single page or screen. To allow for this and still satisfy the construct of fixed texts, pagination is used for texts rather than scrolling. Texts that cover more than one page are presented in their entirety before the student sees the first question.

IT skills: Just as paper-based assessments rely on a set of fundamental skills for working with printed materials, so computer-based assessments rely on a set of fundamental information and communications technology (ICT) skills for using computers. These include knowledge of basic hardware (e.g. keyboard and mouse) and basic conventions (e.g. arrows to move forward and specific buttons to press to execute commands). The intention is to keep such skills to a minimal core level in the tablet based assessment.

Examples of items for addressing the extended PISA-D reading framework

The following six items illustrate the types of questions that can be asked of students at Level 1A and below.

Sample item 1

Directions: Circle **YES** if the sentence makes sense. Circle **NO** if the sentence does not make sense.

The red car had a flat tire.	**YES**	**NO**
Airplanes are made of dogs.	**YES**	**NO**
The happy student read the book last night.	**YES**	**NO**
If the cat had stayed out all night, it would not have been in the house at 2 a.m.	**YES**	**NO**
The man who is taller than the woman and the boy is shorter than both of them.	**YES**	**NO**

Sample item 1 assesses sentence processing tasks and likely corresponds to proficiency level 1c. In PISA-D, sentence processing tasks also included some short sentences with 3 options, with instructions to choose the word that makes the sentence make sense.

Sample item 2

Passage comprehension

In items assessing passage comprehension, respondents are asked to read a passage in which they are required at certain points to select the word that makes sense from the two alternatives provided.

To the editor: Yesterday, it was announced that the cost of riding the bus will increase. The price will go up by twenty percent starting next wife / month. As someone who rides the bus every day, I am upset by this foot / increase. I understand that the cost of gasoline / student has risen. I also understand that riders have to pay a fair price / snake for bus service. I am willing to pay a little more because I rely on the bus to get to object / work. But an increase / uncle of twenty percent is too much.

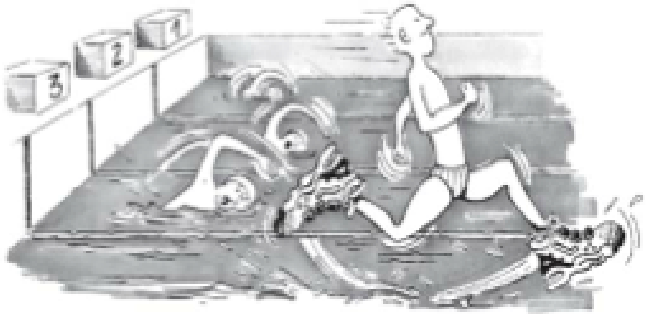
This increase is especially difficult to accept when you see the city's plans to build a new sports stadium. The government will spend millions on this project even though we already have a science / stadium. If we delay the stadium, some of that money can be used to offset the increase in bus fares / views. Then, in a few years, we can decide if we really do need a new sports cloth / arena. Please let the city council know you care about this issue by attending the next public meeting / frames.

Sample item 2 assesses passage comprehension and likely corresponds to proficiency level 1c. In PISA-D, the passage comprehension paragraphs (part of Reading Components) have been modified to have 3 options instead of 2.

Sample item 3

Feel good in your runners

For 14 years the Sports Medicine Centre of Lyon (France) has been studying the injuries of young sports players and sports professionals. The study has established that the best course is prevention … and good shoes.



Knocks, falls, wear and tear...

Eighteen per cent of sports players aged 8 to 12 already have heel injuries. The cartilage of a footballer's ankle does not respond well to shocks, and 25% of professionals have discovered for themselves that it is an especially weak point. The cartilage of the delicate knee joint can also be irreparably damaged and if care is not taken right from childhood (10–12 years of age), this can cause premature osteoarthritis. The hip does not escape damage either and, particularly when tired, players run the risk of fractures as a result of falls or collisions.

According to the study, footballers who have been playing for more than ten years have bony outgrowths either on the tibia or on the heel. This is what is known as "footballer's foot", a deformity caused by shoes with soles and ankle parts that are too flexible.

Protect, support, stabilise, absorb

If a shoe is too rigid, it restricts movement. If it is too flexible, it increases the risk of injuries and sprains. A good sports shoe should meet four criteria:

Firstly, it must **provide exterior protection**: resisting knocks from the ball or another player, coping with unevenness in the ground, and keeping the foot warm and dry even when it is freezing cold and raining.

It must **support the foot**, and in particular the ankle joint, to avoid sprains, swelling and other problems, which may even affect the knee.

It must also provide players with good **stability** so that they do not slip on a wet ground or skid on a surface that is too dry.

Finally, it must **absorb shocks**, especially those suffered by volleyball and basketball players who are constantly jumping.

Dry feet

To avoid minor but painful conditions such as blisters or even splits or athlete's foot (fungal infections), the shoe must allow evaporation of perspiration and must prevent outside dampness from getting in. The ideal material for this is leather, which can be water-proofed to prevent the shoe from getting soaked the first time it rains.

Source: Revue ID (16) 1-15 June 1997.

QUESTION 7.2

According to the article, why should sports shoes not be too rigid?

....

QUESTION 7.4

Look at this sentence from near the end of the article. It is presented here in two parts:

"To avoid minor but painful conditions such as blisters or even splits or athlete's foot (fungal infections),..."	(first part)
"...the shoe must allow evaporation of perspiration and must prevent outside dampness from getting in."	(second part)

What is the relationship between the first and second parts of the sentence?

The second part

A. contradicts the first part.

B. repeats the first part.

C. illustrates the problem described in the first part.

D. gives the solution to the problem described in the first part.

Sample item 3 is a released PISA item that shows a basic informational text. Question 7.2 assesses a student's literal comprehension from the text. Because of the amount of text, students must read, it likely corresponds to level 1a. Question 7.4, on the other hand, assesses integration and interpretation of information and thus is at a higher level of proficiency.

Sample item 4

The Moreland Library System gives new library members a bookmark showing its Hours of Opening. Refer to the bookmark to answer the questions which follow.

Moreland Library System

HOURS OF OPENING *Effective from February 1 1998*

	Brunswick Library	Campbell Turnbull Library	Coburg Library	Fawkner Library	Glenroy Library
Sunday	1pm-5pm	Closed	2pm-5pm	Closed	2pm-5pm
Monday	11am-8pm	11am-5.30pm	1pm-8pm	11am-5.30pm	10am-5.30pm
Tuesday	11am-8pm	11am-8pm	11am-8pm	11am-8pm	10am-8pm
Wednesday	11am-8pm	11am-5pm	10am-8pm	11am-5pm	10am-8pm
Thursday	11am-8pm	11am-5.30pm	10am-8pm	11am-5.30pm	10am-8pm
Friday	11am-5pm	11am-5pm	10am-8pm	11am-5pm	10am-5.30pm
Saturday	10am-1pm	10am-1pm	9am-1pm	10am-1pm	9am-1pm

QUESTION 12.1

What time does the Fawkner Library close on Wednesday?

....

QUESTION 12.2

Which library is still open at 6 p.m. on Friday evening?

A. Brunswick Library

B. Campbell Turnbull Library

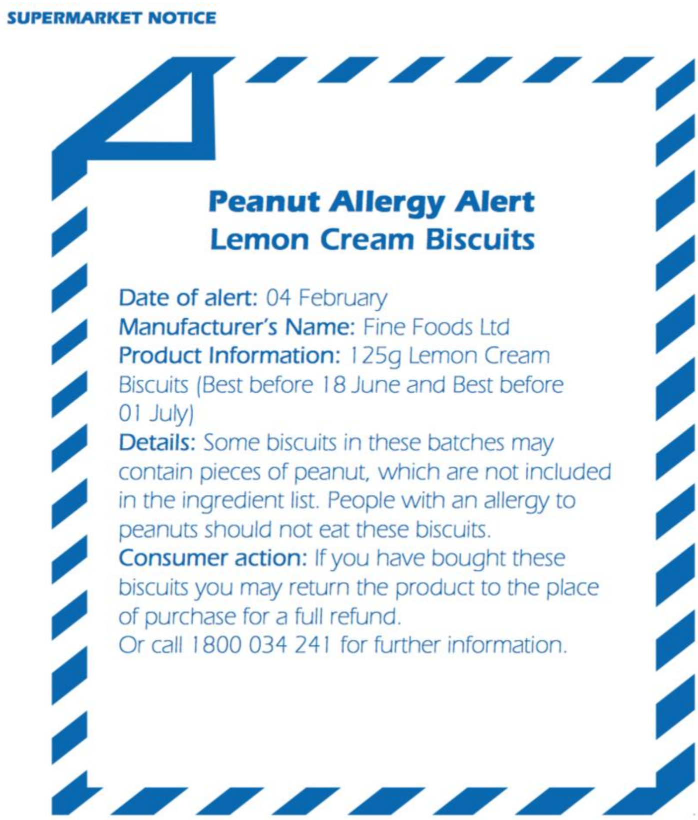
C. Coburg Library

D. Fawkner Library

E. Glenroy Library

Sample item 4 is a released PISA item that assesses basic access and retrieve tasks in a simple non-continuous text. Question 12.1 requires accessing the information directly from a row in the table which is likely to be level 1a or 1b, while question 12.2 requires combining multiple criteria in order to access the correct information which is more likely level 2.

Sample item 5



Question 3: SUPERMARKET NOTICE

What is the name of the company that made the biscuits?

....

Sample item 5 is a released PISA item that assesses basic access and retrieve processes. Question 3, "What is the name of the company that makes the biscuits?" requires a small inference since the text says "manufacturer" rather than "company". Thus, as it stands, the item would likely be at Level 1b of proficiency. However, if it were modified to as "What is the name of the manufacturer that makes the biscuits?", then the item would require a literal match and would be considered at a Level 1c.

NOTES

1. The term processes from the 2018 framework is used instead of the term aspects that is used in previous versions.

2. In 2015 PISA was moved to computer-based delivery with additional consequences for the classification of text types. For more details the reader is referred to the PISA 2015 reading framework.

REFERENCES

Abadzi, H. (2003). Improving adult literacy outcomes: Lessons from cognitive research for developing countries. Washington, DC: The World Bank.

Anderson, R. C., and Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), Comprehension and teaching (pp. 77-117). Newark, DE: International Reading Association.

Baer, J., Kutner, M., and Sabatini, J. (2009). Basic reading skills and the literacy of America's least literate adults: Results from the 2003 National Assessment of Adult Literacy (NAAL) Supplemental Studies (NCES 2009-481). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Binkley, M., Rust, K., and Williams, T. (Eds.). (1997). Reading literacy in an international perspective. Washington, DC: U.S. Department of Education.

Britt, M.A., Goldman, S.R., and Rouet, J.-F. (Eds.). (2012). Reading: From words to multiple texts. New York, NY: Routledge.

Britt, M. A., and Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. In M. J. Lawson and J. R. Kirby (Eds.), The quality of learning: Dispositions, instruction, and mental structures. New York, NY: Cambridge University Press.

Coiro, J., Knobel, M., Lankshear, C., and Leu, D. (2008). Central issues in new literacies and new literacies research. In J. Coiro, M. Knobel, C. Lankshear, and D. Leu (Eds.), The handbook of research on new literacies. Mahwah, NJ: Erlbaum.

Coulombe, S., Tremblay, J-F., and Marchand, S. (2004). Literacy scores, human capital, and growth across fourteen OECD countries. Ottawa, Canada: Statistics Canada.

Council of Europe (1996). Modern languages: learning, teaching, assessment: A common European framework of reference. CC-LANG (95) 5 Rev. IV. Strasbourg, France: Author.

Cunningham, A. E., and Stanovich, K. E. (1998). Early reading acquisition and its relation to reading experience and ability 10 years later. Developmental Psychology 33, 934-945.

Curtis, M. E. (1980). Development of components of reading skill. Journal of Educational Psychology, 72, 656-669.

Dechant, E. (1991). Understanding and teaching reading: An interactive model. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Halpern, D. F. (1989). Thought and knowledge: An introduction to critical thinking. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hirsch, E. D., Jr. (2003). Reading comprehension requires knowledge—of words and the world. American Educator, 27, 10-31.

Holloway, J. H. (1999). Improving the reading skills of adolescents, Educational Leadership, 57(2), 80-82.

Hoover, W.A., and Tunmer, W.E. (1993). The components of reading. In G. G. Thompson, W. E. Tunmer, and T. Nicholson (Eds.), Reading acquisition processes (pp. 1-19). Philadelphia, PA: Multilingual Matters.

Hubbard, R. (1989). Notes from the underground: Unofficial literacy in one sixth grade. Anthropology and Education Quarterly, 20, 291-307.

International Reading Association (2012). Adolescent literacy. Document retrieved from www.literacyworldwide.org, August 28, 2015.

International Telecommunications Union. (2014). Key 2005-2014 ICT data for the world, by geographic regions and by level of development [Excel file]. Geneva (Switzerland): ITU. Retrieved from www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2014.aspx

Kintsch, W. (1998). Comprehension: A paradigm for cognition. Cambridge, MA: Cambridge University Press.

Kirsch, I. (2001). The International Adult Literacy Survey: Understanding what was measured. Princeton, NJ: Educational Testing Service.

Kirsch, I., and Mosenthal, P. B. (1990). Exploring document literacy: Variables underlying the performance of young adults. Reading Research Quarterly, 25(1), 5-30.

McCutchen, D., Green, L., and Abbott, R.D. (2008). Children's morphological knowledge: Links to literacy. Reading Psychology, 29(4), 289–314.

McNamara, D. S., and Magliano, J. (2009). Toward a comprehensive model of comprehension. Psychology of Learning and Motivation, 51, 297-384.

Nagy, W., and Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. Reading Research Quarterly, 47, 91-108.

Oakhill, J. V., Cain, K., and Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. Language and Cognitive Processes, 18, 443-468.

Ouellet, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. Journal of Educational Psychology, 98, 554–566.

Perfetti, C. A. (1985). Reading ability. New York, NY: Oxford University Press.

Perfetti, C. A. (2003). The universal grammar of reading. Scientific Studies of Reading, 7(1), 3-24.

Perfetti, C. A., Landi, N., and Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling and C. Hulme (Eds.), The science of reading: A handbook (pp. 227–247). Malden, MA: Blackwell.

Pew Internet and American Life Project (2005). Trends 2005: Internet: The mainstreaming of online life. Washington, DC.

Rand Corporation. (2002). Reading for understanding: Toward an RandD program in reading comprehension. Santa Monica, CA: Author.

Rayner, K., and Reichle, E. D. (2010). Models of the reading process. Wiley Interdisciplinary Reviews. Cognitive Science, 1(6), 787–799. doi:10.1002/wcs.68

Reinking, D. (1994). Electronic literacy. Perspectives in Reading Research, 4.

Richter, T., and Rapp, D. N. (2014). Comprehension and validation of text information: Introduction to the special issue. Discourse Processes, 51, 1-6.

Rumelhart, D. E. (1985). Toward an interactive model of reading. In H. Singer and R. B. Ruddell (Eds.), Theoretical models and the processes of reading (3rd ed.). Newark, DE: International.

Sabatini, J. P., and Bruce, K. M. (2009). Reading components: A conceptual framework, OECD Working Paper No. 33. Retrieved from: www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=enandcote=edu/wkp(2009)12

Shanahan, T., and Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. Harvard Educational Review, 78(1), 40–59.

Shetzer, H. and Warschauer, M. (2000). An electronic literacy approach to network-based language teaching. In M. Warschauer and R. Kem (Eds.), Network-based language teaching: Concepts and practice (pp. 175-185). New York: Cambridge University Press.

Smith, M. C., Mikulecky, L., Kibby, M. W., and Dreher, M. J. (2000). What will be the demands of literacy in the workplace in the next millennium? Reading Research Quarterly, 35(3), 378-383.

Sticht, T. G. (Ed.). (1975). Reading for working: A functional literacy anthology. Alexandria, VA: Human Resources Research Organization.

Stiggins, R. J. (1982). An analysis of the dimensions of job-related reading. Reading World, 82, 237-247.

Stine-Morrow, E. A. L., Miller, L. M. S., and Hertzog, C. (2006). Aging and self-regulated language processing. Psychological Bulletin, 132, 582-606.

Sweets, R., and Meates, A. (2004), ICT and low achievers: What does PISA tell us? Budapest, Hungary, and Paris, France: Hungarian Ministry of Education.

Van den Broek, P., Virtue, S., Everson, M., Tzeng, Y., and Sung, Y. (2002). Comprehension and memory of science texts: Inferential processes and the construction of a mental representation. In J. Otero, J. Leon, and A. C. Graesser (Eds.), The psychology of science text comprehension (pp. 131-154). Mahwah, NJ: Erlbaum.

Warschauer, M. (1999), Electronic literacies: Language culture and power in online education. Mahwah, NJ: Lawrence Erlbaum Associates.

Werlich, E. (1976). A text grammar of English. Heidelberg, Germany: Quelle.

Zwaan, R. A., and Singer, M. (2003). Text comprehension. In A. C. Graesser, M. A. Gernsbacher, and S. R. Goldman (Eds.), Handbook of discourse processes (pp. 83-122). Mahwah, NJ: Erlbaum.

CHAPTER 3. PISA FOR DEVELOPMENT MATHEMATICS FRAMEWORK

This chapter defines "mathematical literacy" as assessed in the Programme for International Student Assessment (PISA) and the extensions to the PISA mathematics framework that have been designed for the PISA for Development (PISA-D) project. It explains the processes, content knowledge and contexts reflected in PISA-D's mathematics problems, and provides several sample items. The chapter also discusses how student performance in mathematics is measured and reported.

What is new in PISA-D? Extensions to the PISA mathematical literacy framework

The objective of the PISA-D mathematics framework is to extend the PISA framework to be able to measure mathematical skills of students who perform at or below the lowest level on the standard math PISA scale. The outcomes of such measurement should provide reliable data that could help to plan the most effective ways of improving those students' mathematical skills. This extended framework is applicable not only for students but also for 14-16-years-olds that are out of school or not enrolled in PISA's target grades (grade 7 or above).

To achieve this objective of the framework, it would seem quite natural to concentrate attention on some very basic "numeracy skills" such as fluency in performing simple arithmetical operations. However, it would not be an effective solution. While certainly some of these skills are needed to perform at the highest levels of PISA – such as arithmetical fluency, understanding of basic mathematical concepts, being able to recognise and identify graphs, and understanding of math vocabulary – they are not the focus of PISA.

The implementation of the PISA measurement of mathematics was preceded by a scientific discussion on the role of teaching mathematics. The references section of this chapter lists the most important scientific publications behind those discussions. The framework itself provides a comprehensive explanation of the most important conclusions, culminating in the concept of mathematical literacy. In short, it stresses the primary importance of the ability to use mathematics in a wide variety of contexts. The international success of PISA confirms that this is the widely accepted way of understanding the primary goal of learning mathematics in today's world.

From this perspective, mastering the most basic technical skills is not enough. Of course it is important to be able to perform arithmetical operations, but performing these operations it is not sufficient to get by mathematically in real life. To put this knowledge to use, one necessarily needs at least the basic skills of choosing the right model and selecting a strategy or an explanation. These skills constitute the core of the PISA understanding of mathematical literacy.

Identifying some of even the most basic technical skills as a measure of mathematical competencies would be, in this context, quite misleading. It could direct the attention of the users of the PISA results toward those skills as the primary education target. However, this approach would give little chance to their students to become more mathematically literate.

The PISA-D mathematics framework adheres to the core idea of mathematical literacy, as defined by PISA. Therefore it is designed just as an extension to the PISA 2015 mathematics framework. The framework is therefore designed for the measurement of essentially the same basic skills as PISA 2015. The extensions aim at expanding coverage at the lower ability levels essentially in two ways: by using more straightforward, simply formulated items and by suggesting a very careful analysis of students' attempts to solve the problem. However, the items used in PISA-D will also test the ability to choose the right model and select a strategy or an explanation. Only in this way will PISA-D have a chance to prove useful in improving the mathematical literacy of the students.

The extensions made to the PISA 2015 framework in PISA-D are an attempt to gain more information about students that currently perform below Level 1. In the mathematics framework, these extensions occur in three locations: descriptions of the proficiencies, where proficiency level 1 was renamed as 1a and two new proficiency levels were added, 1b and 1c; adding five new activities to the process descriptors; and adding four new skills to the table relating the mathematical processes to the fundamental mathematical capabilities.

The PISA 2015 framework (OECD, 2016) continues the description and illustration of the PISA mathematics assessment as set out in the 2012 framework, when mathematics was re-examined and updated for use as the major domain in that cycle.

For PISA 2015, the computer was the primary mode of delivery for all domains, including mathematical literacy. The 2015 framework was updated to reflect the change in delivery mode, and includes a discussion of the considerations of transposing paper items to a screen and examples of what the results look like. The definition and constructs of mathematical literacy however, remain unchanged and consistent with those used in 2012. It is important to note that PISA-D includes a paper-based test for the in-school population and a tablet-based test for the out-of-school population. For this reason, therefore, the sections in this chapter dealing with computer-based assessment of mathematics only apply to the out-of-school assessment.

This chapter is organised into three major sections. The first section, "Defining Mathematical Literacy," explains the theoretical underpinnings of the PISA mathematics assessment, including the formal definition of the mathematical literacy construct. The second section, "Organising the domain of mathematics," describes three aspects: a) the mathematical *processes* and the *fundamental mathematical capabilities* (in previous frameworks the "competencies") underlying those processes; b) the way mathematical *content* knowledge is organised in the PISA 2015 framework, and the content knowledge that is relevant to an assessment of 15-year-old students; and c) the *contexts* in which students will face mathematical challenges. The third section, "Assessing Mathematical Literacy", outlines the approach taken to apply the elements of the framework previously described, including the response formats, item scoring, reporting proficiency, testing mathematical literacy among the out-of-school population and examples of items for addressing the extended PISA-D framework.

Defining mathematical literacy

An understanding of mathematics is central to a young person's preparedness for life in modern society. A growing proportion of problems and situations encountered in daily life, including in professional contexts, require some level of understanding of mathematics, mathematical reasoning and mathematical tools, before they can be fully understood and addressed. Mathematics is a critical tool for young people as they confront issues and challenges in personal, occupational, societal, and scientific aspects of their lives. It is thus important to have an understanding of the degree to which young people emerging from school are adequately prepared to apply mathematics to understanding important issues and solving meaningful problems. An assessment at age 15 provides an early indication of how individuals may respond in later life to the diverse array of situations they will encounter that involve mathematics.

The construct of mathematical literacy used in this chapter is based on PISA 2015 and is intended to describe the capacities of individuals to reason mathematically and use mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. This conception of mathematical literacy supports the importance of students developing a strong understanding of concepts of pure mathematics and the benefits of being engaged in explorations in the abstract world of mathematics. The construct of mathematical literacy, as defined for PISA, strongly emphasises the need to develop students' capacity to use mathematics in context, and it is important that they have rich experiences in their mathematics classrooms to accomplish this. For PISA 2012, mathematical literacy was defined as shown in Box 3.1. This is the definition that is used in the PISA 2015 and PISA-D assessment also.

Box 3.1 The PISA 2015 definition of mathematical literacy

> Mathematical literacy is an individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens.

The focus of the language in the definition of mathematical literacy is on active engagement in mathematics, and is intended to encompass reasoning mathematically and using mathematical concepts, procedures, facts and tools in describing, explaining and predicting phenomena. In particular, the verbs 'formulate,' 'employ,' and 'interpret' point to the three processes in which students as active problem solvers will engage.

The language of the definition was also intended to integrate the notion of mathematical modelling, which has historically been a cornerstone of the PISA framework for mathematics (e.g. OECD, 2004), into the PISA 2015 definition of mathematical literacy. As individuals use mathematics and mathematical tools to solve problems in contexts, their work progresses through a series of stages (individually developed later in the document).

The modelling cycle is a central aspect of the PISA conception of students as active problem solvers; however, it is often not necessary to engage in every stage of the modelling cycle, especially in the context of an assessment (Niss et al., 2007). The problem solver frequently carries out some steps of the modelling cycle but not all of them, (e.g. when using graphs), or goes around the cycle several times to modify earlier decisions and assumptions.

The definition also acknowledges that mathematical literacy helps individuals to recognise the role that mathematics plays in the world and in helping they make the kinds of well-founded judgements and decisions required of constructive, engaged and reflective citizens.

Mathematical tools mentioned in the definition refer to a variety of physical and digital equipment, software and calculation devices. The 2015 computer-based survey as well as the PISA-D tablet based test included an online calculator as part of the computer-based test material provided for some questions.

Organising the domain of mathematics

The PISA mathematics framework defines the domain of mathematics for the PISA survey and describes an approach to the assessment of the mathematical literacy of 15-year-olds. That is, PISA assesses the extent to which 15-year-old students can handle mathematics adeptly when confronted with situations and problems – the majority of which are presented in real-world contexts.

For purposes of the assessment, the PISA 2015 definition of mathematical literacy can be analysed in terms of three interrelated aspects:

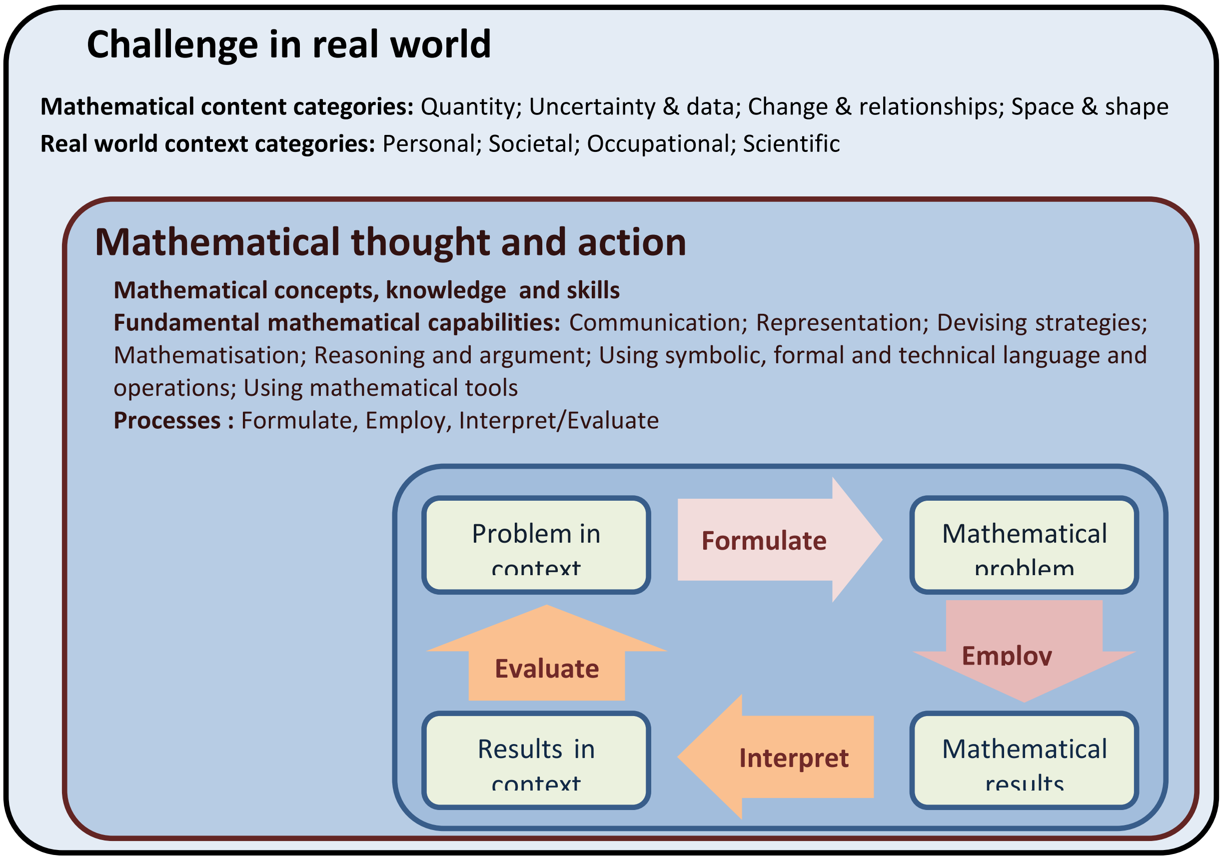
- The mathematical processes that describe what individuals do to connect the context of the problem with mathematics and thus solve the problem, and the capabilities that underlie those processes;
- The mathematical content that is targeted for use in the assessment items; and
- The contexts in which the assessment items are located.

The following sections elaborate these aspects. In highlighting these aspects of the domain, the PISA 2012 mathematics framework, which was also used in PISA 2015 and PISA-D, helps ensure that assessment items developed for the survey reflect a range of processes, content, and contexts, so that, considered as a whole, the set of assessment items effectively operationalises what this framework defines as mathematical literacy. To illustrate the aspects of mathematical literacy, examples are available in the PISA 2012 Assessment and Analytical Framework (OECD, 2013) and on the PISA website (www.oecd.org/pisa/).

Several questions, based on the PISA 2015 definition of mathematical literacy, lie behind the organisation of this section of the chapter. They are:

- What processes do individuals engage in when solving contextual mathematical problems, and what capabilities do we expect individuals to be able to demonstrate as their mathematical literacy grows?
- What mathematical content knowledge can we expect of individuals – and of 15-year-old students in particular?
- In what contexts can mathematical literacy able to be observed and assessed?

Figure 3.1 A model of mathematical literacy in practice



Mathematical processes and the underlying mathematical capabilities

Mathematical processes

The definition of mathematical literacy refers to an individual's capacity to formulate, employ, and interpret mathematics. These three words, formulate, employ and interpret, provide a useful and meaningful structure for organising the mathematical processes that describe what individuals do to connect the context of a problem with the mathematics and thus solve the problem. Items in the PISA 2015 and the PISA-D mathematics survey will be assigned to one of three mathematical processes:

- formulating situations mathematically
- employing mathematical concepts, facts, procedures, and reasoning
- interpreting, applying and evaluating mathematical outcomes.

It is important for both policy makers and those engaged more closely in the day-to-day education of students to know how effectively students are able to engage in each of these processes. The *formulating* process indicates how effectively students are able to recognise and identify opportunities to use mathematics in problem situations and then provide the necessary mathematical structure needed to formulate that contextualised problem into a mathematical form. The *employing* process indicates how well students are able to perform computations and manipulations and apply the concepts and facts that they know to arrive at a mathematical solution to a problem formulated mathematically. The *interpreting* process indicates how effectively students are able to reflect upon mathematical solutions or conclusions, interpret them in the context of a real-world problem, and determine whether the results or conclusions are reasonable. Students' facility at applying mathematics to problems and situations is dependent on skills inherent in all three of these processes, and an understanding of their effectiveness in each category can help inform both policy-level discussions and decisions being made closer to the classroom level.

In an effort to better measure the capabilities of Level 1b and 1c students, specific extensions have been made in PISA-D to the descriptions of the processes formulate, employ and interpret. The five additions are intended to better describe students' attempts to apply mathematical processes. The approach taken acknowledges that before students may be fully capable of utilising processes, they must first be able to identify and select an appropriate model, strategy or argument.

Formulating situations mathematically

The word *formulate* in the mathematical literacy definition refers to individuals being able to recognise and identify opportunities to use mathematics and then provide mathematical structure to a problem presented in some contextualised form. In the process of formulating situations mathematically, individuals determine where they can extract the essential mathematics to analyse, set up, and solve the problem. They translate from a real-world setting to the domain of mathematics and provide the real-world problem with mathematical structure, representations, and specificity. They reason about and make sense of constraints and assumptions in the problem. Specifically, this process of formulating situations mathematically includes activities such as the following:

- identifying the mathematical aspects of a problem situated in a real-world context and identifying the significant variables
- recognising mathematical structure (including regularities, relationships and patterns) in problems or situations

- simplifying a situation or problem in order to make it amenable to mathematical analysis
- identifying constraints and assumptions behind any mathematical modelling and simplifications gleaned from the context
- representing a situation mathematically, using appropriate variables, symbols, diagrams and standard models
- representing a problem in a different way, including organising it according to mathematical concepts and making appropriate assumptions
- understanding and explaining the relationships between the context-specific language of a problem and the symbolic and formal language needed to represent it mathematically
- translating a problem into mathematical language or a representation
- recognising aspects of a problem that correspond with known problems or mathematical concepts, facts or procedures
- using technology (such as a spreadsheet or the list facility on a graphing calculator) to portray a mathematical relationship inherent in a contextualised problem.

In addition to the activities listed above, the following activity has been added to PISA-D:

- selecting an appropriate model from a list.

Employing mathematical concepts, facts, procedures and reasoning

The word *employ* in the mathematical literacy definition refers to individuals being able to apply mathematical concepts, facts, procedures, and reasoning to solve mathematically formulated problems to obtain mathematical conclusions. In the process of employing mathematical concepts, facts, procedures, and reasoning to solve problems, individuals perform the mathematical procedures needed to derive results and find a mathematical solution (e.g. performing arithmetic computations, solving equations, making logical deductions from mathematical assumptions, performing symbolic manipulations, extracting mathematical information from tables and graphs, representing and manipulating shapes in space, and analysing data). They work on a model of the problem situation, establish regularities, identify connections between mathematical entities and create mathematical arguments. Specifically, this process of employing mathematical concepts, facts, procedures and reasoning includes activities such as:

- devising and implementing strategies for finding mathematical solutions
- using mathematical tools, including technology, to help find exact or approximate solutions
- applying mathematical facts, rules, algorithms and structures when finding solutions
- manipulating numbers, graphical and statistical data and information, algebraic expressions and equations, and geometric representations
- making mathematical diagrams, graphs and constructions and extracting mathematical information from them

- using and switching between different representations in the process of finding solutions
- making generalisations based on the results of applying mathematical procedures to find solutions
- reflecting on mathematical arguments and explaining and justifying mathematical results.

In addition to the activities listed above, the following activities have been added to PISA-D:

- performing a simple calculation
- drawing a simple conclusion
- selecting an appropriate strategy from a list.

Interpreting, applying and evaluating mathematical outcomes

The word *interpret* used in the mathematical literacy definition focuses on the abilities of individuals to reflect upon mathematical solutions, results, or conclusions and interpret them in the context of real-life problems. This involves translating mathematical solutions or reasoning back into the context of a problem and determining whether the results are reasonable and make sense in the context of the problem. This mathematical process category encompasses both the "interpret" and "evaluate" arrows noted in the previously defined model of mathematical literacy in practice (see Figure 3.1). Individuals engaged in this process may be called upon to construct and communicate explanations and arguments in the context of the problem, reflecting on both the modelling process and its results. Specifically, this process of interpreting, applying, and evaluating mathematical outcomes includes activities such as:

- interpreting a mathematical result back into the real-world context
- evaluating the reasonableness of a mathematical solution in the context of a real-world problem
- understanding how the real-world impacts the outcomes and calculations of a mathematical procedure or model in order to make contextual judgments about how the results should be adjusted or applied
- explaining why a mathematical result or conclusion does, or does not, make sense given the context of a problem
- understanding the extent and limits of mathematical concepts and mathematical solutions
- critiquing and identifying the limits of the model used to solve a problem.

In addition to the activities listed above, the following activity has been added to PISA-D:

- evaluating a mathematical outcome in terms of the context.

Desired distribution of items by mathematical process

The goal in constructing the assessment is to achieve a balance that provides approximately equal weighting between the two processes that involve making a connection between the real world and the mathematical world and the process that calls for students to be able to work on a mathematically formulated problem. Table 3.1 shows the desired distribution of items by process for PISA 2015 and PISA-D (both for the in- and out-of-school instruments).

Table 3.1 Desired distribution of mathematics items by process category

Process category	Percentage of items in PISA 2015	Percentage of items in PISA-D
Formulating situations mathematically	25	25
Employing mathematical concepts, facts, procedures and reasoning	50	50
Interpreting, applying and evaluating mathematical outcomes	25	25
Total	**100**	**100**

The desired distribution specifies the blueprint for the selection of items according to important aspects of the domain frameworks. Item selection is based on the assessment design as well as item characteristics related to a number of framework aspects, including process, content and context category, as well as consideration of the items' psychometric properties and appropriateness for this assessment. Following the assessment, the actual distributions of items across the framework aspects will be described in relation to the desired distributions. The extent to which the item pool for the assessment meets the framework specifications will be discussed in the technical report in the context of practical constraints in the item selection process.

Fundamental mathematical capabilities underlying the mathematical processes

A decade of experience in developing PISA items and analysing the ways in which students respond to items has revealed that there is a set of fundamental mathematical capabilities that underpins each of these reported processes and mathematical literacy in practice. The work of Mogens Niss and his Danish colleagues (Niss, 2003; Niss and Jensen, 2002; Niss and Højgaard, 2011) identified eight capabilities – referred to as "competencies" by Niss and in the 2003 framework (OECD, 2003) – that are instrumental to mathematical behaviour.

The PISA 2015 and PISA-D framework uses a modified formulation of this set of capabilities, which condenses the number from eight to seven based on an investigation of the operation of the competencies through previously administered PISA items (Turner et al., 2013). There is wide recognition of the need to identify such a set of general mathematical capabilities to complement the role of specific mathematical content knowledge in mathematics learning. Prominent examples include the eight mathematical practices of the Common Core State Standards Initiative in the United States (2010), the four key processes (representing, analysing, interpreting and evaluating, and communicating and reflecting) of England's Mathematics National Curriculum (Qualifications and Curriculum Authority, 2007), and the process standards in the National Council of Teachers of Mathematics Principles and Standards for School Mathematics (NCTM, 2000). These cognitive capabilities are available to or learnable by individuals in order to understand and engage with the world in a mathematical way, or to solve problems. As the level of mathematical literacy possessed by an individual increases, that individual is able to draw to an increasing degree on the fundamental mathematical capabilities (Turner and Adams, 2012). Thus, increasing activation of fundamental mathematical capabilities is associated with increasing item difficulty. This observation has been used as the basis of the descriptions of different proficiency levels of mathematical literacy reported in previous PISA surveys and discussed later in this framework.

The seven fundamental mathematical capabilities used in the PISA 2015 and PISA-D framework are as follows:

- *Communication*: Mathematical literacy involves communication. The individual perceives the existence of some challenge and is stimulated to recognise and understand a problem situation. Reading, decoding, and interpreting statements, questions, tasks or objects enables the individual to form a mental model of the situation, which is an important step in understanding, clarifying, and formulating a problem. During the solution process, intermediate results may need to be summarised and presented. Later on, once a solution has been found, the problem solver may need to present the solution, and perhaps an explanation or justification, to others.

- *Mathematising*: Mathematical literacy can involve transforming a problem defined in the real world to a strictly mathematical form (which can include structuring, conceptualising, making assumptions and/or formulating a model), or interpreting or evaluating a mathematical outcome or a mathematical model in relation to the original problem. The term mathematising is used to describe the fundamental mathematical activities involved.

- *Representation*: Mathematical literacy very frequently involves representations of mathematical objects and situations. This can entail selecting, interpreting, translating between, and using a variety of representations to capture a situation, interact with a problem, or to present one's work. The representations referred to include graphs, tables, diagrams, pictures, equations, formulae and concrete materials.

- *Reasoning and argument:* A mathematical ability that is called on throughout the different stages and activities associated with mathematical literacy is referred to as reasoning and argument. This capability involves logically rooted thought processes that explore and link problem elements so as to make inferences from them, check a justification that is given, or provide a justification of statements or solutions to problems.

- *Devising strategies for solving problems*: Mathematical literacy frequently requires devising strategies for solving problems mathematically. This involves a set of critical control processes that guide an individual to effectively recognise, formulate, and solve problems. This skill is characterised as selecting or devising a plan or strategy to use mathematics to solve problems arising from a task or context, as well as guiding its implementation. This mathematical capability can be demanded at any of the stages of the problem solving process.

- *Using symbolic, formal and technical language and operations*: Mathematical literacy requires using symbolic, formal, and technical language and operations. This involves understanding, interpreting, manipulating, and making use of symbolic expressions within a mathematical context (including arithmetic expressions and operations) governed by mathematical conventions and rules. It also involves understanding and utilising formal constructs based on definitions, rules, and formal systems and also using algorithms with these entities. The symbols, rules and systems used will vary according to what particular mathematical content knowledge is needed for a specific task to formulate, solve or interpret the mathematics.

- *Using mathematical tools:*[1] The final mathematical capability that underpins mathematical literacy in practice is using mathematical tools. Mathematical tools encompass physical tools such as measuring instruments, as well as calculators and computer-based tools that are becoming more widely available. This ability involves knowing about and being able to make use of various tools that may assist mathematical activity, and knowing about the limitations of such tools. Mathematical tools can also have an important role in communicating results.

These capabilities are evident to varying degrees in each of the three mathematical processes. The ways in which these capabilities manifest themselves within the three processes are described in Figure 3.2.

A good guide to the empirical difficulty of items can be obtained by considering which aspects of the fundamental mathematical capabilities are required for planning and executing a solution (Turner, 2012, Turner and Adams, 2012; Turner et al., 2013). The easiest items will require the activation of few capabilities and in a relatively straightforward way. The hardest items require complex activation of several capabilities. Predicting difficulty requires consideration of both the number of capabilities and the complexity of activation required. Based on the modifications to the proficiencies and processes for PISA-D, it was necessary to add particular skills to support these modifications. Four skills were added to the table in order to provide better understanding of the extensions of the mathematical process descriptions. These skills also support the capabilities delineated in the proficiencies 1b and 1c.

Figure 3.2 Relationship between mathematical processes (top horizontal row) and fundamental mathematical capabilities (left-most vertical column)

	Formulating situations mathematically	**Employing mathematical concepts, facts, procedures and reasoning**	**Interpreting, applying and evaluating mathematical outcomes**
Communicating	Read, decode, and make sense of statements, questions, tasks, objects or images, in order to form a mental model of the situation	Articulate a solution, show the work involved in reaching a solution, and/or summarise and present intermediate mathematical results	Construct and communicate explanations and arguments in the context of the problem
Mathematising	Identify the underlying mathematical variables and structures in the real-world problem, and make assumptions so that they can be used. For PISA-D, select a model appropriate to the context of real-world problems has been added.	Use an understanding of the context to guide or expedite the mathematical solving process, e.g. working to a context-appropriate level of accuracy	Understand the extent and limits of a mathematical solution that are a consequence of the mathematical model employed
Representation	Create a mathematical representation of real-world information. For PISA-D, select a representation appropriate to the context has been added.	Make sense of, relate, and use a variety of representations when interacting with a problem	Interpret mathematical outcomes in a variety of formats in relation to a situation or use; compare or evaluate two or more representations in relation to a situation
Reasoning and argument	Explain, defend or provide a justification for the identified or devised representation of a real-world situation	Explain, defend, or provide a justification for the processes and procedures used to determine a mathematical result or solution Connect pieces of information to arrive at a mathematical solution, make generalisations, or create a multi-step argument. For PISA-D, select an appropriate justification has been added.	Reflect on mathematical solutions and create explanations and arguments that support, refute or qualify a mathematical solution to a contextualised problem
Devising strategies for solving problems	Select or devise a plan or strategy to mathematically reframe contextualised problems	Activate effective and sustained control mechanisms across a multi-step procedure leading to a mathematical solution, conclusion, or generalisation	Devise and implement a strategy in order to interpret, evaluate and validate a mathematical solution to a contextualised problem. For PISA-D, implement a given strategy has been added.

Using symbolic, formal, and technical language and operations	Use appropriate variables, symbols, diagrams and standard models in order to represent a real-world problem using symbolic/formal language	Understand and utilise formal constructs based on definitions, rules, and formal systems as well as employing algorithms	Understand the relationship between the context of the problem and representation of the mathematical solution. Use this understanding to help interpret the solution in context and gauge the feasibility and possible limitations of the solution.
Using mathematical tools	Use mathematical tools in order to recognise mathematical structures or to portray mathematical relationships	Know about and be able to make appropriate use of various tools that may assist in implementing processes and procedures for determining mathematical solutions	Use mathematical tools to ascertain the reasonableness of a mathematical solution and any limits and constraints on that solution, given the context of the problem

Mathematical content knowledge

An understanding of mathematical content – and the ability to apply that knowledge to the solution of meaningful contextualised problems – is important for citizens in the modern world. That is, to solve problems and interpret situations in personal, occupational, societal and scientific contexts, there is a need to draw upon certain mathematical knowledge and understandings.

Mathematical structures have been developed over time as a means to understand and interpret natural and social phenomena. In schools, the mathematics curriculum is typically organised around content strands (e.g. number, algebra and geometry) and detailed topic lists that reflect historically well-established branches of mathematics and that help in defining a structured curriculum. However, outside the mathematics classroom, a challenge or situation that arises is usually not accompanied by a set of rules and prescriptions that shows how the challenge can be met. Rather it typically requires some creative thought in seeing the possibilities of bringing mathematics to bear on the situation and in formulating it mathematically. Often a situation can be addressed in different ways, drawing on different mathematical concepts, procedures, facts or tools.

Since the goal of PISA is to assess mathematical literacy, an organisational structure for mathematical content knowledge has been developed based on the mathematical phenomena that underlie broad classes of problems and which have motivated the development of specific mathematical concepts and procedures. Because national mathematics curricula are typically designed to equip students with knowledge and skills that address these same underlying mathematical phenomena, the outcome is that the range of content arising from organising content this way is closely aligned with that typically found in national mathematics curricula. This framework lists some content topics appropriate for assessing the mathematical literacy of 15-year-old students based on analyses of national standards from eleven countries.

To organise the domain of mathematics for purposes of assessing mathematical literacy, it is important to select a structure that grows out of historical developments in mathematics, that encompasses sufficient variety and depth to reveal the essentials of mathematics, and that also represents, or includes, the conventional mathematical strands in an acceptable way. Thus, a set of content categories that reflects the range of underlying mathematical phenomena was selected for the PISA 2015 framework, consistent with the categories used for previous PISA surveys.

The following list of content categories, therefore, will be used in PISA 2015 and PISA-D to meet the requirements of historical development, coverage of the domain of mathematics and the underlying phenomena that motivate its development, and reflection of the major strands of school curricula. These four categories characterise the range of mathematical content that is central to the discipline and illustrate the broad areas of content used in the test items for PISA 2015 and PISA-D:

- *Change and relationships*
- *Space and shape*
- *Quantity*
- *Uncertainty and data*

With these four categories, the mathematical domain can be organised in a way that ensures a spread of items across the domain and focuses on important mathematical phenomena, but at the same time, avoids a too fine division that would work against a focus on rich and challenging mathematical problems based on real situations. While categorisation by content category is important for item development and selection, and for reporting of assessment results, it is important to note that some specific content topics may materialise in more than one content category. For example, a released PISA item called *Pizzas* involves determining which of two round pizzas, with different diameters and different costs but the same thickness, is the better value (see the "Illustrative PISA-D mathematics items" section to view this item and an analysis of its attributes). This item draws on several areas of mathematics, including measurement, quantification (value for money, proportional reasoning and arithmetic calculations), and change and relationships (in terms of relationships among the variables and how relevant properties change from the smaller pizza to the larger one.) This item was ultimately categorised as a *change and relationships* item since the key to the problem lies in students being able to relate the change in areas of the two pizzas (given a change in diameter) and a corresponding change of price. Clearly, a different item involving circle area might be classified as a *space and shape* item. Connections between aspects of content that span these four content categories contribute to the coherence of mathematics as a discipline and are apparent in some of the assessment items selected for the PISA 2015 assessment.

The broad mathematical content categories and the more specific content topics appropriate for 15-year-old students described later in this section reflect the level and breadth of content that is eligible for inclusion on the PISA 2015 and PISA-D survey. Narrative descriptions of each content category and the relevance of each to solving meaningful problems are provided first, followed by more specific definitions of the kinds of content that are appropriate for inclusion in an assessment of mathematical literacy of 15-year-old students and out-of-school youth. These specific topics reflect commonalities found in the expectations set by a range of countries and educational jurisdictions. The standards examined to identify these content topics are viewed as evidence not only of what is taught in mathematics classrooms in these countries but also as indicators of what countries view as important knowledge and skills for preparing students of this age to become constructive, engaged and reflective citizens.

Descriptions of the mathematical content knowledge that characterise each of the four categories — *change and relationships*, *space and shape*, *quantity*, and *uncertainty and data* — are provided below.

Change and relationships

The natural and designed worlds display a multitude of temporary and permanent relationships among objects and circumstances, where changes occur within systems of interrelated objects or in circumstances where the elements influence one another. In many cases these changes occur over time, and in other cases changes in one object or quantity are related to changes in another. Some of these situations involve discrete change; others change continuously. Some relationships are of a permanent, or invariant, nature. Being more literate about change and relationships involves understanding fundamental types of change and recognising when they occur in order to use suitable mathematical models to describe and predict change. Mathematically this means modelling the change and the relationships with appropriate functions and equations, as well as creating, interpreting, and translating among symbolic and graphical representations of relationships.

Change and relationships is evident in such diverse settings as growth of organisms, music, and the cycle of seasons, weather patterns, employment levels, and economic conditions. Aspects of the traditional mathematical content of functions and algebra, including algebraic expressions, equations and inequalities, and tabular and graphical representations, are central in describing, modelling, and interpreting change phenomena. For example, the released PISA unit *Walking* (see the "Illustrative PISA-D mathematics items" section) contains two items that exemplify the *change and relationships* category since the focus is on the algebraic relationships between two variables, requiring students to activate their algebraic knowledge and skills. Students are required to employ a given formula for pacelength – a formula expressed in algebraic form – to determine pacelength in one item and walking speed in the other. Representations of data and relationships described using statistics also are often used to portray and interpret change and relationships, and a firm grounding in the basics of number and units is also essential to defining and interpreting *change and relationships*. Some interesting relationships arise from geometric measurement, such as the way that changes in perimeter of a family of shapes might relate to changes in area, or the relationships among lengths of the sides of triangles.

Space and shape

Space and shape encompasses a wide range of phenomena that are encountered everywhere in our visual and physical world: patterns, properties of objects, positions and orientations, representations of objects, decoding and encoding of visual information, and navigation and dynamic interaction with real shapes as well as with representations. Geometry serves as an essential foundation for *space and shape*, but the category extends beyond traditional geometry in content, meaning, and method, drawing on elements of other mathematical areas such as spatial visualisation, measurement, and algebra. For instance, shapes can change, and a point can move along a locus, thus requiring function concepts. Measurement formulas are central in this area. The manipulation and interpretation of shapes in settings that call for tools ranging from dynamic geometry software to global positioning system (GPS) software are included in this content category.

PISA assumes that the understanding of a set of core concepts and skills is important to mathematical literacy relative to *space and shape*. Mathematical literacy in the area of *space and shape* involves a range of activities such as understanding perspective (for example in paintings), creating and reading maps, transforming shapes with and without technology, interpreting views of three-dimensional scenes from various perspectives and constructing representations of shapes.

Quantity

The notion of *quantity* may be the most pervasive and essential mathematical aspect of engaging with, and functioning in, our world. It incorporates the quantification of attributes of objects, relationships, situations and entities in the world, understanding various representations of those quantifications, and judging interpretations and arguments based on quantity. To engage with the quantification of the world involves understanding measurements, counts, magnitudes, units, indicators, relative size, and numerical trends and patterns. Aspects of quantitative reasoning—such as number sense, multiple representations of numbers, elegance in computation, mental calculation, estimation, and assessment of reasonableness of results—are the essence of mathematical literacy relative to *quantity*.

Quantification is a primary method for describing and measuring a vast set of attributes of aspects of the world. It allows for the modelling of situations, for the examination of change and relationships, for the description and manipulation of space and shape, for organising and interpreting data, and for the measurement and assessment of uncertainty. Thus mathematical literacy in the area of *Quantity* applies knowledge of number and number operations in a wide variety of settings.

Uncertainty and data

In science, technology and everyday life, uncertainty is a given. Uncertainty is therefore a phenomenon at the heart of the mathematical analysis of many problem situations, and the theory of probability and statistics as well as techniques of data representation and description have been established to deal with it. The *uncertainty and data* content category includes recognising the place of variation in processes, having a sense of the quantification of that variation, acknowledging uncertainty and error in measurement, and knowing about chance. It also includes forming, interpreting, and evaluating conclusions drawn in situations where uncertainty is central. The presentation and interpretation of data are key concepts in this category (Moore, 1997).

There is uncertainty in scientific predictions, poll results, weather forecasts, and economic models. There is variation in manufacturing processes, test scores, and survey findings, and chance is fundamental to many recreational activities enjoyed by individuals. The traditional curricular areas of probability and statistics provide formal means of describing, modelling, and interpreting a certain class of uncertainty phenomena, and for making inferences. In addition, knowledge of number and of aspects of algebra such as graphs and symbolic representation contribute to facility in engaging in problems in this content category. The focus on the interpretation and presentation of data is an important aspect of the Uncertainty and data category.

Desired distribution of items by content category

The desired distribution of items selected for PISA 2015 and for PISA-D (both in- and out-of-school instruments) across the four content categories is shown in Table 3.2. The goal in constructing the survey is a balanced distribution of items with respect to content category, since all of these domains are important for constructive, engaged and reflective citizens.

Table 3.2 Desired distribution of mathematics items by content category

Content category	Percentage of items in PISA 2015	Percentage of items in PISA-D
Change and relationships	25	25
Space and shape	25	25
Quantity	25	25
Uncertainty and data	25	25
Total	**100**	**100**

Content topics for guiding the assessment of mathematical literacy for 15-year-old students

To effectively understand and solve contextualised problems involving *change and relationships, space and shape, quantity* and *uncertainty and data* requires drawing upon a variety of mathematical concepts, procedures, facts, and tools at an appropriate level of depth and sophistication. As an assessment of mathematical literacy, PISA strives to assess the levels and types of mathematics that are appropriate for 15-year-old students on a trajectory to become constructive, engaged and reflective citizens able to make well-founded judgments and decisions. It is also the case that PISA, while not designed or intended to be a curriculum-driven assessment, strives to reflect the mathematics that students have likely had the opportunity to learn by the time they are 15 years old.

The content included in PISA-D and PISA 2015 is the same as that developed in PISA 2012.The four content categories of *change and relationships*, *space and shape*, *quantity* and *uncertainty and data* serve as the foundation for identifying this range of content, yet there is not a one-to-one mapping of content topics to these categories. For example, proportional reasoning comes into play in such varied contexts as making measurement conversions, analysing linear relationships, calculating probabilities and examining the lengths of sides in similar shapes. The following content is intended to reflect the centrality of many of these concepts to all four content categories and reinforce the coherence of mathematics as a discipline. It intends to be illustrative of the content topics included in PISA-D, rather than an exhaustive listing:

- *Functions*: the concept of function, emphasising but not limited to linear functions, their properties, and a variety of descriptions and representations of them. Commonly used representations are verbal, symbolic, tabular, and graphical.

- *Algebraic expressions:* verbal interpretation of and manipulation with algebraic expressions, involving numbers, symbols, arithmetic operations, powers, and simple roots.

- *Equations and inequalities:* linear and related equations and inequalities, simple second-degree equations, and analytic and non-analytic solution methods

- *Co-ordinate systems:* representation and description of data, position, and relationships.

- *Relationships within and among geometrical objects in two and three dimensions:* Static relationships such as algebraic connections among elements of figures (e.g. the Pythagorean theorem as defining the relationship between the lengths of the sides of a right triangle), relative position, similarity and congruence, and dynamic relationships involving transformation and motion of objects, as well as correspondences between two- and three-dimensional objects.

- *Measurement:* Quantification of features of and among shapes and objects, such as angle measures, distance, length, perimeter, circumference, area, and volume.

- *Numbers and units:* Concepts; representations of numbers and number systems, including properties of integer and rational numbers; relevant aspects of irrational numbers; as well as quantities and units referring to phenomena such as time, money, weight, temperature, distance, area and volume, and derived quantities and their numerical description.

- *Arithmetic operations:* the nature and properties of these operations and related notational conventions.

- *Percents, ratios and proportions:* numerical description of relative magnitude and the application of proportions and proportional reasoning to solve problems.

- *Counting principle*s: Simple combinations and permutations.

- *Estimation:* Purpose-driven approximation of quantities and numerical expressions, including significant digits and rounding.

- *Data collection, representation and interpretation:* nature, genesis, and collection of various types of data, and the different ways to represent and interpret them.

- *Data variability and its description:* Concepts such as variability, distribution, and central tendency of data sets, and ways to describe and interpret these in quantitative terms.

- *Samples and sampling:* Concepts of sampling and sampling from data populations, including simple inferences based on properties of samples.

- *Chance and probability:* notion of random events, random variation and its representation, chance and frequency of events, and basic aspects of the concept of probability.

Contexts

The choice of appropriate mathematical strategies and representations is often dependent on the context in which a problem arises. Being able to work within a context is widely appreciated to place additional demands on the problem solver (see Watson and Callingham, 2003, for findings about statistics). For the PISA survey, it is important that a wide variety of contexts are used. This offers the possibility of connecting with the broadest possible range of individual interests and with the range of situations in which individuals operate in the 21st century.

For purposes of the PISA-D mathematics framework, four context categories have been defined and are used to classify assessment items developed for the PISA survey:

- *Personal* – Problems classified in the personal context category focus on activities of one's self, one's family, or one's peer group. The kinds of contexts that may be considered personal include (but are not limited to) those involving food preparation, shopping, games, personal health, personal transportation, sports, travel, personal scheduling, and personal finance.

- *Occupational* – Problems classified in the occupational context category are centred on the world of work. items categorised as occupational may involve (but are not limited to) such things as measuring, costing and ordering materials for building, payroll/accounting, quality control, scheduling/inventory, design/architecture and job-related decision making. Occupational contexts may relate to any level of the workforce, from unskilled work to the highest levels of professional work, although items in the PISA survey must be accessible to 15-year-old students.

- *Societal* – Problems classified in the societal context category focus on one's community (whether local, national, or global). They may involve (but are not limited to) such things as voting systems, public transport, government, public policies, demographics, advertising, national statistics, and economics. Although individuals are involved in all of these things in a personal way, in the societal context category, the focus of problems is on the community perspective.

- *Scientific* – Problems classified in the scientific category relate to the application of mathematics to the natural world and issues and topics related to science and technology. Particular contexts might include (but are not limited to) such areas as weather or climate, ecology, medicine, space science, genetics, measurement and the world of mathematics itself. Items that are intramathematical, where all the elements involved belong in the world of mathematics, fall within the scientific context.

PISA items were arranged in units that share stimulus material. It is therefore usually the case that all items in the same unit belonged to the same context category. Exceptions do arise; for example stimulus material may be examined from a personal point of view in one item and a societal point of view in another. When an item involved only mathematical constructs without reference to the contextual elements of the unit within which it is located, it was allocated to the context category of the unit. In the unusual case of a unit involving only mathematical constructs and being without reference to any context outside of mathematics, the unit was assigned to the scientific context category.

Using these context categories provided the basis for selecting a mix of item contexts and ensures that the assessment reflects a broad range of uses of mathematics, ranging from everyday personal uses to the scientific demands of global problems. Moreover it was important that each context category be populated with assessment items having a broad range of item difficulties. Given that the major purpose of these context categories is to challenge students in a broad range of problem contexts, each category was designed to contribute substantially to the measurement of mathematical literacy. It should not be the case that the difficulty level of assessment items representing one context category is systematically higher or lower than the difficulty level of assessment items in another category.

In identifying contexts that may be relevant, it is critical to keep in mind that a purpose of the assessment is to gauge the use of mathematical content knowledge, processes, and capabilities that students have acquired by age 15. Contexts for assessment items, therefore, were selected in light of relevance to students' interests and lives, and the demands that will be placed upon them as they enter society as constructive, engaged, and reflective citizens. National Project Managers from countries participating in the PISA-D survey were involved in judging the degree of such relevance.

Desired distribution of items by context category

The desired distribution of items selected for PISA 2015 and for PISA-D (both in- and out-of-school instruments) across the four content categories is shown in Table 3.3. With this balanced distribution, no single context type is allowed to dominate, providing students with items that span a broad range of individual interests and a range of situations that they might expect to encounter in their lives.

Table 3.3 Desired distribution of mathematics items by context category

Context category	Percentage of items in PISA 2015	Percentage of items in PISA-D
Personal	25	25
Occupational	25	25
Societal	25	25
Scientific	25	25
Total	**100**	**100**

Assessing mathematical literacy

This section outlines the approach taken to apply the elements of the framework described in previous sections to PISA 2015 and PISA-D. This includes the structure of the mathematics component of the PISA-D survey, arrangements for transferring the paper-based trend items to a computer-based delivery, and reporting mathematical proficiency.

Response formats

Three types of response format are used to assess mathematical literacy in PISA 2015 and PISA-D: open constructed-response, closed constructed-response and selected-response (simple and complex multiple-choice) items. Open constructed-response items require a somewhat extended written response from a student. Such items also may ask the student to show the steps taken or to explain how the answer was reached. These items require trained experts to manually code student responses.

Closed constructed-response items provide a more structured setting for presenting problem solutions, and they produce a student response that can be easily judged to be either correct or incorrect. Often student responses to questions of this type can be keyed into data capture software, and coded automatically, but some must be manually coded by trained experts. The most frequent closed constructed-responses are single numbers.

Selected-response items require students to choose one or more responses from a number of response options. Responses to these questions can usually be automatically processed. About equal numbers of each of these response formats is used to construct the survey instruments.

Item scoring

Although the majority of the items were dichotomously scored (that is, responses are awarded either credit or no credit), the open constructed-response items can sometimes involve partial credit scoring, which allows responses to be assigned credit according to differing degrees of "correctness" of responses. For each such item, a detailed coding guide that allows for full credit, partial credit, or no credit was provided to persons trained in the coding of student responses across the range of participating countries to ensure coding of responses was done in a consistent and reliable way. To maximise the comparability between the paper-based and computer-based assessment, careful attention is given to the scoring guides in order to ensure that the important elements were included.

Reporting proficiency in mathematics

The outcomes of the PISA mathematics survey are reported in a number of ways. Estimates of overall mathematical proficiency are obtained for sampled students in each participating country, and a number of proficiency levels are defined. Descriptions of the degree of mathematical literacy typical of students in each level are also developed. For PISA 2003, scales based on the four broad content categories were developed. In Figure 3.3, descriptions for the six proficiency levels reported for the overall PISA mathematics scale in 2012 are presented. These form the basis for the PISA 2015 mathematics scale and the PISA-D mathematics scale. The finalised 2012 scale is used to report the PISA 2015 outcomes. For PISA-D, in addition, the existing Level 1 was renamed to Level 1a, and the table describing the proficiencies has been extended to include levels 1b and 1c.

Fundamental mathematical capabilities play a central role in defining what it means to be at different levels of the scales for mathematical literacy overall and for each of the reported processes. For example, in the proficiency scale description for Level 4 (see Figure 3.3), the second sentence highlights aspects of mathematising and representation thatare evident at this level. The final sentence highlights the characteristic communication, reasoning and argument of Level 4, providing a contrast with the short communications and lack of argument of Level 3 and the additional reflection of Level 5. In an earlier section of this framework and in Figure 3.2, each of the mathematical processes was described in terms of the fundamental mathematical capabilities that individuals might activate when engaging in that process.

Figure 3.3 Summary description of the eight levels of mathematics proficiency in PISA-D

Level	Lower score limit	Percentage of students across OECD countries at each proficiency level in PISA 2015	Percentage of students across 23 middle- and low-income countries at each proficiency level in PISA 2015	Descriptor
6	669	2.3%	0.3%	At Level 6, students can conceptualise, generalise and utilise information based on their investigations and modelling of complex problem situations, and can use their knowledge in relatively non-standard contexts. They can link different information sources and representations and flexibly translate among them. Students at this level are capable of advanced mathematical thinking and reasoning. These students can apply this insight and understanding, along with a mastery of symbolic and formal mathematical operations and relationships, to develop new approaches and strategies for attacking novel situations. Students at this level can reflect on their actions, and can formulate and precisely communicate their actions and reflections regarding their findings, interpretations, arguments, and the appropriateness of these to the original situation.
5	607	8.4%	1.5%	At Level 5, students can develop and work with models for complex situations, identifying constraints and specifying assumptions. They can select, compare and evaluate appropriate problem-solving strategies for dealing with complex problems related to these models. Students at this level can work strategically using broad, well-developed thinking and reasoning skills, appropriate linked representations, symbolic and formal characterisations, and insight pertaining to these situations. They begin to reflect on their work and can formulate and communicate their interpretations and reasoning.
4	545	18.6%	5.3%	At Level 4, students can work effectively with explicit models for complex concrete situations that may involve constraints or call for making assumptions. They can select and integrate different representations, including symbolic, linking them directly to aspects of real-world situations. Students at this level can utilise their limited range of skills and can reason with some insight, in straightforward contexts. They can construct and communicate explanations and arguments based on their interpretations, arguments and actions.
3	482	24.8%	12.6%	At Level 3, students can execute clearly described procedures, including those that require sequential decisions. Their interpretations are sufficiently sound to be a base for building a simple model or for selecting and applying simple problem- solving strategies. Students at this level can interpret and use representations based on different information sources and reason directly from them. They typically show some ability to handle percentages, fractions and decimal numbers, and to work with proportional relationships. Their solutions reflect that they have engaged in basic interpretation and reasoning.
2	420	22.5%	21.6%	At Level 2, students can interpret and recognise situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formulae, procedures or conventions to solve problems involving whole numbers. They are capable of making literal interpretations of the
1a	358	14.9%	26.3%	At Level 1a students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are almost always obvious and follow immediately from the given stimuli.
1b	*	8.5% (percentage of students scoring below Level 1 in PISA 2015)	32.4% (percentage of students scoring below Level 1 in PISA 2015)	At Level 1b students can understand questions involving everyday contexts where all relevant information is clearly given and defined in a short syntactically simple text. They are able to follow clearly prescribed instructions. They can perform the first step of a two-step solution of a problem.
1c	*			At Level 1c students can understand questions involving simple, everyday contexts where all relevant information is clearly given and defined in a very short syntactically simple text. They are able to follow a single clearly prescribed instruction. They can solve problems limited to a single step or operation.

Note: * Will be available after the main study.

In order to gain useful information for these new Levels 1b and 1c, it is vital that context and language not interfere with the mathematics being assessed. To this end, the context and language must be carefully considered.

The context for both 1b and 1c should be situations that students encounter on a daily basis. Examples of these contexts may include money, temperature, food, time, date, weight, size and distance. All items should be concrete and not abstract. The focus of the item should be mathematical only. The understanding of the context should not interfere with the performance of the item.

Equally important is to have all items formulated in the simplest possible terms. Sentences should be short and direct. Compound sentences, compound nouns and conditional sentences should be avoided. Vocabulary used in the items must be carefully examined to ensure that students will have a clear understanding of what is being required. In addition, special care will be given to ensure that no extra difficulty is added due to a heavy text load or by a context that is unfamiliar to students based on their cultural background.

Items designed for Level 1c should just ask for a single step or operation. However, it is important to note that a single step or operation is not limited to an arithmetical step. This step might be demonstrated by making a selection or identifying some information. All parts of the modelling cycle can and should be used to measure mathematical abilities of students at Levels 1b and 1c.

Testing mathematical literacy among the out-of-school population

For the out-of-school population, item selection focused on the scale at or below Level 2 with an emphasis on the lower end of the scale in terms of the distribution of items. The selection process was similar to that for the in-school population: coverage of all processes was maintained and contexts of the items were reviewed to ensure appropriateness for what individuals would encounter in an out-of-school context.

Box 3.2 Delivery mode

The PISA-D school based assessment is paper based, while the out-of-school assessment is conducted on a tablet computer. To ensure comparability between the tests, the tablet based instruments for PISA-D are formed by a subgroup of the items used for the paper based assessment. All of these items were originally designed for a paper based assessment, so when moving to a tablet based delivery, care was taken to maintain comparability between the assessments. The PISA 2015 framework describes some factors that must be considered when transposing items from paper to computer mode. These elements were also taken into account when designing the out-of-school instrument for PISA-D.

Item types: The computer provides a range of opportunities for designers of test items, including new item formats (e.g. drag-and-drop, hotspots). Since the PISA-D tablet bases tests uses a sub group of items form the paper based test, there is less opportunity to exploit innovative item types and the majority of response formats remains unchanged.

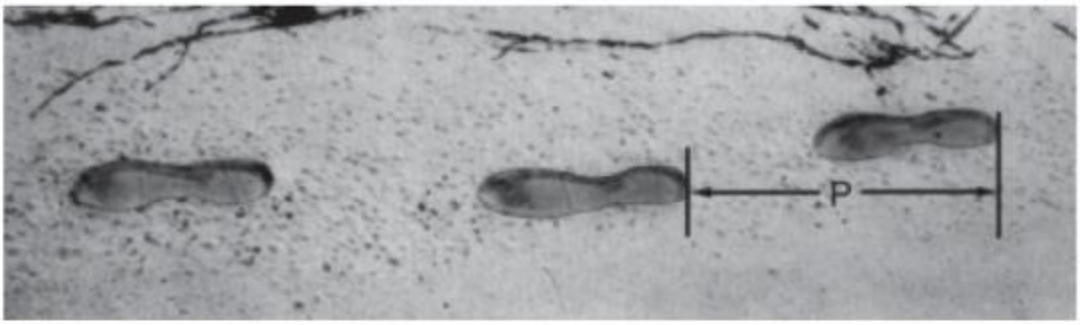
Stimulus presentation: A feature of fixed texts defined in the construct is that “the extent or amount of the text is immediately visible to the reader”. Clearly, it is impossible, both on paper and on a screen, to have long texts displayed on a single page or screen. To allow for this and still satisfy the construct of fixed texts, pagination is used for texts rather than scrolling. Texts that cover more than one page are presented in their entirety before the student sees the first question.

IT skills: Just as paper-based assessments rely on a set of fundamental skills for working with printed materials, so computer-based assessments rely on a set of fundamental information and communications technology (ICT) skills for using computers. These include knowledge of basic hardware (e.g. keyboard and mouse) and basic conventions (e.g. arrows to move forward and specific buttons to press to execute commands). The intention is to keep such skills to a minimal core level in the tablet based assessment.

Examples of items for addressing the extended PISA-D mathematics framework

The following items illustrate proficiency Levels 1a, 1b and 1c. In each case the tasks involved are described and an explanation is given about why the item fits a certain proficiency level.

Sample item 1



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

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

n = number of steps per minute, and

P = pacelength in metres.

Heiko has a pacelength that is 0.5 meters. Using this formula, how many steps per minute, *n*, does Heiko take each minute?

For this item, the student simply needs to substitute the values into the formula and solve. Since the pacelength is given (0.5 meters), the student only needs to complete a single operational step. Multiplying both sides of the equation by 0.5 gives a value for *n* of 70. This addresses the added process "performing a simple calculation." A student who only substitutes the values correctly would meet the requirements of proficiency 1c, while a student who also performs the operational step correctly would meet the requirements of proficiency 1b.

Sample item 2

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

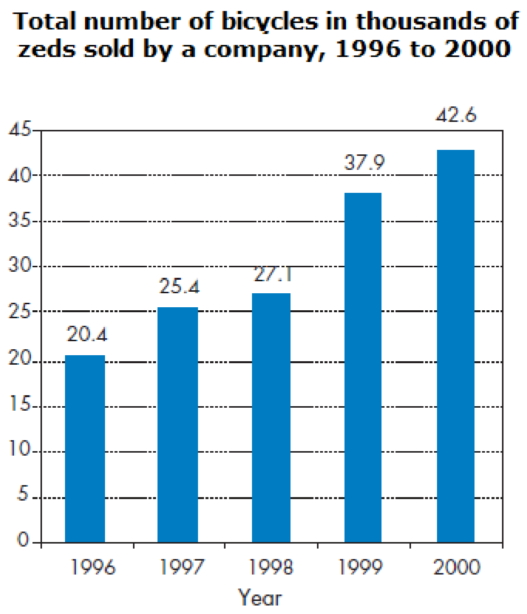
1 SGD = 4.2 ZAR

Mei-Ling changed 3000 Singapore dollars into South African rand at this exchange rate. Choose a correct method from those listed. Then calculate *n*, the amount of South African rand Mei-Ling received after the exchange.

$\frac{1}{4.2} = \frac{n}{3000}$ $\quad\quad$ $\frac{1}{3000} = \frac{4.2}{n}$ $\quad\quad$ $4.2n = 3000$ $\quad\quad$ $n = 3000(4.2)$

For this item, the student is given 4 methods to solve for *n*. Two of these methods will result in a correct value for *n*. The expectation is that a student will be able to select one of the correct methods and then solve for the value of *n*. This addresses the added process "selecting an appropriate model from a list." If a student is able to choose one of the correct methods, the requirements for proficiency 1c are met. If a student is also able to solve for *n* correctly, the requirements for proficiency 1b are met.

Sample item 2



What was the total number of bicycles in thousands of zeds that were sold by the company in 1998?

For this item, the student is given a graph and asked to answer a question about it. This addresses the second part of the process "making mathematical diagrams, graphs, and constructions and extracting mathematical information from them." The student performs one, single step by retrieving information from the graph. This would satisfy the requirements of proficiency 1c.

Sample item 3

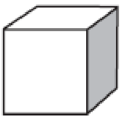
Nick wants to pave the rectangular patio of his new house with bricks. The patio has a length of 5.25 meters and a width of 3.00 meters. One box of bricks can pave 2 square meters.

Calculate the number of boxes of bricks Nick needs for the entire patio.

For this item, the student must perform two steps to arrive at a correct solution. First the student must find the area of the patio. Then for the second step, the student must divide the number of square meters by 2 in order to find the total number of boxes of bricks. This addresses multiple processes in understanding what must be done, devising a strategy, and performing the calculations. By awarding partial credit for a student who only successfully finds the area, proficiency 1b is addressed. Proficiency 1a is addressed if a student does both steps correctly.

Sample item 4

Susan likes to build blocks using small cubes like the one shown in the following diagram:



Small cube

Susan will build a block as shown in Diagram A below:

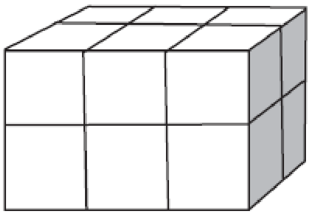


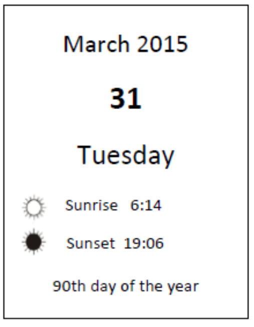
Diagram A

How many small cubes will Susan need to build the block shown in Diagram A?

For this item, the student needs to determine the number of small cubes needed to build the larger block. In doing this, the process "applying mathematical facts, rules, algorithms, and structures when finding solutions" is addressed. Since this is a simple, one-step problem, it meets the requirements of proficiency 1c.

Sample item 5

The picture represents one page of a calendar.



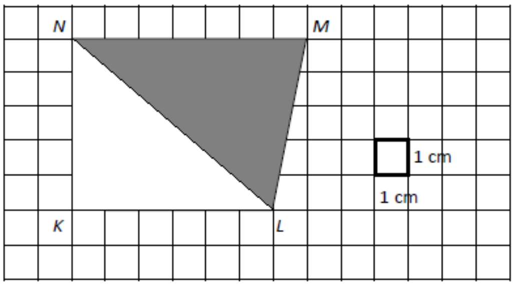
How much time will pass on that day from sunrise until sunset?

A. 12 hours and 52 minutes
B. 13 hours and 8 minutes
C. 13 hours and 32 minutes
D. 13 hours and 52 minutes

For this item, the student must determine the elapsed time. To solve this successfully, students must recognize the modification to the normal subtraction algorithm when regrouping. Because this recognition is required, even though it is one single operational step, there is a thought process that keeps this from being a 1b or 1c. This is more of a traditional 1a item.

Sample item 6

A quadrilateral has been divided into two triangles: white triangle *KLN* and grey triangle *LMN*.



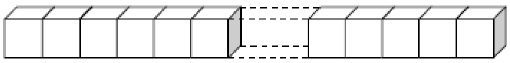
Which of the following statements are true? Mark T, when the statement is true or F – when it is false.

Area of the gray triangle is larger than the area of the white one.	T	F
The areas of those triangles differ by 5 cm^2.	T	F

For this item, all the information is provided by the picture. The first question is at proficiency Level 1c as it tests the understanding the concept of an area. Students who can only deal with area when a formula is involved would have difficulty with this item, as it focuses more on reasoning and recognition. No formula is necessary. The second question, still at proficiency Level 1c, goes a little deeper and tests the understanding of a unit of area. Again, those who understand the concept of area will count squares; those who understand area only through a formula will have difficulty.

Sample item 7

A cube of volume 1 m^3 has been cut off into small cubes of edge length 1 cm. If we put those small cubes one after another, as shown in the picture, we would get a square prism.



Which of the following statements are true? Mark T, when the statement is true or F – when it is false.

The volume of this square prism is 100 times larger than the volume of the original cube.	T	F
One of the edges of this square prism has length 10 km.	T	F

For this item, the student must demonstrate an understanding of the concept of volume. It requires no calculations at all, only simple reasoning. It is a proficiency Level 1b item, because there is modelling hidden here. We do not see the large cube in the picture. The second question requires a calculation. Students who go by “common sense” usually choose the wrong answer. The student has to ignore his or her intuitive judgment and pick the mathematical way of dealing with the problem. This item is proficiency Level 1b.

NOTES

1. In some countries, “mathematical tools” can also refer to established mathematical procedures such as algorithms. For the purposes of the PISA framework, “mathematical tools” refers only to the physical and digital tools described in this section.

REFERENCES

Bennett, R.E. (2003), *Online Assessment and the Comparability of Score Meaning*, Research Memorandum, Educational Testing Service, Princeton, NJ.

Bennett, R.E., J. Braswell, A. Oranje, B. Sandene, B. Kaplan and F. Yan (2008), "Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP", *Journal of Technology, Learning, and Assessment*, Vol. 6/9.

Common Core State Standards Initiative (2010), *Common Core State Standards for Mathematics, Common Core State Standards initiative*, Washington, DC, www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.

Devlin, K. (1994), *Mathematics: The science of patterns: The search for order in life, mind and the universe*, W.H. Freeman Scientific American Library, Washington, DC.

Educational Testing Service (2008), *Online Assessment in Mathematics and Writing: Reports from the NAEP technology-based assessment project*, Educational Testing Service, Princeton, NJ.

Hoyles, C., A. Wolf, S. Molyneux-Hodgson and P. Kent (2002), *Mathematical Skills in the Workplace: Final report to the Science Technology and Mathematics Council, Project Report*, Institute of Education, University of London, Science, Technology and Mathematics Council, London, http://eprints.ioe.ac.uk/1565/1/Hoyles2002MathematicalSkills.pdf.

Mason, B., M. Patry and D. Bernstein (2001), "An examination of the equivalence between non-adaptive computer based and traditional testing", *Journal of Education Computing Research*, Vol. 1/24, pp. 29-39.

Moore, D. (1997), "New pedagogy and new content: The case of statistics", *International Statistical Review*, Vol. 2/65, pp. 123-137.

National Council of Teachers of Mathematics (2000), *Principles and standards for school mathematics*, National Council of Teachers of Mathematics, Reston, VA, www.nctm.org/standards/.

Niss, M. (2003), "Mathematical competencies and the learning of mathematics: The Danish KoM project", in A. Gagatsis and S. Papastavridis (eds.), *3rd Mediterranean conference on mathematics education*, The Hellenic Mathematical Society and Cyprus Mathematical Society, Athens, pp. 115-124, http://w3.msi.vxu.se/users/hso/aaa_niss.pdf.

Niss, M., W. Blum and P. Galbraith (2007), "Introduction", in W. Blum, P. Galbraith, H.-W. Henn and M. Niss, (eds.) *Modelling and applications in mathematics education* (The 14th ICMI Study), Spring, New York, pp. 3-32.

Niss, M. and T. Højgaard (eds.) (2011), "Competencies and mathematical learning: Ideas and inspiration for the development of mathematics teaching and learning in Denmark", Ministry of Education Report No. 485, Roskilde University, Roskilde, https://pure.au.dk/portal/files/41669781/THJ11_MN_KOM_in_english.pdf.

Niss, M. and T.H. Jensen (2002), "Kompetencer og matematiklæring: Ideer og inspiration til udvikling af matematikundervisning i Danmark, uddannelsesstyrelsens temahæfteserie", No. 18, Ministry of Education, Copenhagen, http://pub.uvm.dk/2002/kom/.

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

OECD (2014), *Technical report of the Survey of Adult Skills (PIAAC)*, pre-publication, OECD, Paris, www.oecd.org/site/piaac/_Technical%20Report_17OCT13.pdf.

OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, PISA, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264190511-en.

OECD (2010), *Pathways to Success: How knowledge and skills at age 15 shape future lives in Canada*, PISA, OECD Publishing, Paris, www.OECD.org/dataOECD/59/35/44574748.pdf.

OECD (2004), *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264101739-en.

Qualifications and Curriculum Authority (2007), *Mathematics: Programme of study for key stage 3 and attainment targets*, Qualifications and Curriculum Authority, London, http://media.education.gov.uk/assets/files/pdf/q/mathematics%202007%20programme%20of%20study%20for%20key%20stage%203.pdf

Richardson, M., J. Baird, J. Ridgway, M. Ripley, D. Shorrocks-Taylor and M. Swan (2002), "Challenging minds? Students' perceptions of computer-based world class tests of problem solving", in *Computers in Human Behavior*, Vol. 18/6, pp. 633-649.

Sandene, B. et al. (2005), *Online Assessment in Mathematics and Writing: Reports from the NAEP Technology-Based Assessment Project, Research and Development Series* (NCES 2005–457), US Department of Education, National Center for Education Statistics, US Government Printing Office, Washington, DC.

Stacey, K. and D. Wiliam (2013), "Technology and assessment in mathematics", in M.A. (Ken) Clements, A. Bishop, C. Keitel, J. Kilpatrick and F. Leung (eds), *Third International Handbook of Mathematics Education*, Spring, New York, pp.721-752.

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

Turner, R. and R.J. Adams (2012) "Some drivers of test item difficulty in mathematics: An analysis of the competency rubric", paper presented at the annual meeting of the American Educational Research Association (AERA), 13-17 April, Vancouver, http://research.acer.edu.au/PISA/7/.

Turner, R., J, Dossey, W. Blum and M. Niss (2013), "Using mathematical competencies to predict item difficulty in PISA", in M. Prenzel, M. Kobarg, K. Schöps and S. Rönnebeck (eds.), *Research on PISA: Research outcomes of the PISA Research Conference 2009*, Springer, New York, pp.23-27.

Watson, J. M. and R. Callingham (2003), "Statistical literacy: A complex hierarchical construct", *Statistics Education Research Journal*, Vol. 2/2, pp. 3-46.

Wang, S., H. Jiao, M.J. Young, T.E. Brooks and J. Olson (2007), "A meta-analysis of testing mode effects in grade K–12 mathematics tests", *Educational and Psychological Measurement*, Vol. 67, pp. 219-238.

CHAPTER 4. PISA FOR DEVELOPMENT SCIENCE FRAMEWORK

This chapter defines "scientific literacy" as assessed in the Programme for International Student Assessment (PISA) and the extensions to the PISA science framework that have been designed for the PISA for Development (PISA-D) project. It describes the types of contexts, knowledge and competencies reflected in PISA-D's science problems, and provides several sample items. The chapter also discusses how student performance in science is measured and reported.

What is new in PISA-D? Extensions to the PISA scientific literacy framework

This chapter presents the assessment framework for science in PISA-D and shows how it specifically addresses the needs and contexts of using PISA for assessing student competency in a wider range of countries. This chapter explains how the PISA 2015 science framework has been extended to provide more information regarding student performance at the lower levels of proficiency. While PISA-D's out-of-school component does not include the science domain, this framework is applicable for students who are in school as well as for 15-year-olds that are out of school.

PISA establishes a baseline level – proficiency Level 2, on a scale with 6 as the highest level and 1b the lowest – at which individuals begin to demonstrate the competencies that will enable them to participate effectively and productively in life as students, workers and citizens. The extensions made to the PISA-D science framework are an attempt to gain more information about students that currently perform at or below Level 1. PISA-D builds on the PISA 2015 science framework, extending it to yet a lower level of performance (1c) to gather precise data on the science skills of the lowest performers. These extensions have been achieved by describing how the expectations of the three competencies – Explain phenomena scientifically, Evaluate and design scientific enquiry, and Interpret data and evidence scientifically – can help distinguish the differences between level 1a, 1b, and 1c students based on increasing, but limited, cognitive demands. In general, all level 1 items make less extensive demands on students' knowledge, and require less cognitive processing. In order to provide greater clarity, the document also explains what kinds of competencies and displays of understanding are not expected.

This chapter adds elements to indicate what it is reasonable to assess and what is expected of students who might perform at Levels 1 and 2 on the PISA scales, suggesting that assessment at these levels should be restricted, wherever possible, to items that make the lowest level of cognitive demand. In addition, to reduce the linguistic demands and cognitive load of any item, careful attention should be paid to simplifying the language of any item and removing extraneous text.

The PISA-D science framework adheres to the core idea of scientific literacy, as defined by PISA. The 1999, 2004 and 2006 PISA frameworks have elaborated a conception of scientific literacy as the central construct for science assessment. These documents have established a broad consensus among science educators of the concept of scientific literacy. The framework for PISA 2015 refines and extends the previous construct, in particular by drawing on the PISA 2006 framework that was used as the basis for assessment in 2006, 2009 and 2012. In 2015 science was the main domain, and PISA-D has no main domains. So those sections that are not relevant to the PISA-D framework – and hence much of the discussion on attitudes – have been omitted or made briefer in this framework.

Scientific literacy matters at both the national and international levels as humanity faces major challenges in providing sufficient water and food, controlling diseases, generating sufficient energy and adapting to climate change (UNEP, 2012). Many of these issues arise, however, at the local level where individuals may be faced with decisions about practices that affect their own health and food supplies, the appropriate use of materials and new technologies, and decisions about energy use. Dealing with all of these challenges will require a major contribution from science and technology. Yet, as argued by the European Commission, the solutions to political and ethical dilemmas involving science and technology "cannot be the subject of informed debate unless young people possess certain scientific awareness" (European Commission, 1995, p.28). Moreover, "this does not mean turning everyone into a scientific expert, but enabling them to fulfil an enlightened role in making choices which affect their environment and to understand in broad terms the social implications of debates between experts" (ibid. p.28). Given that knowledge of science and science-based technology contributes significantly to individuals' personal, social, and professional lives, an understanding of science and technology is thus central to a young person's "preparedness for life".

The concept of scientific literacy in this framework 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. For instance, new scientific knowledge leads to the development of new technologies (think of the advances in material science that led to the development of the transistor in 1948). Likewise, new technologies can lead to new scientific knowledge (think of how knowledge of the universe has been transformed through the development of better telescopes). Individuals make decisions and choices that influence the directions of new technologies (consider the decision to drive a smaller, more fuel-efficient car). Scientifically literate individuals should therefore be able 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. Therefore, individuals need to be able to weigh the potential benefits and risks of applying scientific knowledge to themselves and society.

Scientific literacy also requires not just knowledge of the concepts and theories of science but also knowledge of the common procedures and practices associated with scientific enquiry and how these enable science to advance. Therefore, individuals who are scientifically literate have a knowledge of the major concepts and ideas that form the foundation of scientific and technological thought; how such knowledge has been derived; and the degree to which such knowledge is proved by evidence or theoretical explanations.

Undoubtedly, many of the challenges of the 21st century will require innovative solutions that have a basis in scientific thinking and scientific discovery. Societies will require a cadre of well-educated scientists to undertake the research and nurture the innovation that will be essential to meet the economic, social and environmental challenges that the world faces.

For all of these reasons, scientific literacy is perceived to be a key competency (Rychen and Salganik, 2003) and defined in terms of the ability to use knowledge and information interactively – that is "an understanding of how it [a knowledge of science] changes the way one can interact with the world and how it can be used to accomplish broader goals" (ibid.:10). As such, it represents a major goal for science education for all students. Therefore, the view of scientific literacy that forms the basis for the 2015 international assessment of 15-year-old students is a response to the question: What is important for young people to know, value and be able to do in situations involving science and technology?

This chapter is organised into the following sections. The first section, "Defining scientific literacy," explains the theoretical underpinnings of the PISA science assessment, including the formal definition of the scientific literacy construct and describing the three competencies required for scientific literacy. The second section, "Organising the domain of science," describes the four inter-related aspects that form the definition of scientific literacy: contexts, competencies, knowledge and attitudes. The third section, "Assessing scientific literacy", outlines the approach taken to apply the elements of the framework previously described, including cognitive demand, test characteristics, reporting proficiency, testing scientific literacy among the out-of-school population and examples of items for addressing the extended PISA-D Framework.

Defining scientific literacy

Current thinking about the desired outcomes of science education is rooted strongly in a belief that an understanding of science is so important that it should be a feature of every young person's education (American Association for the Advancement of Science, 1989; Confederacion de Sociedades Cientificas de España, 2011; Fensham, 1985; Millar and Osborne, 1998; National Research Council, 2012 Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005;

Taiwan Ministry of Education, 1999). Indeed, in many countries science is an obligatory element of the school curriculum from kindergarten until the completion of compulsory education.

Many of the documents and policy statements cited above give pre-eminence to an education for citizenship. However, many of the curricula for school science across the world are based on a view that the primary goal of science education should be the preparation of the next generation of scientists (Millar and Osborne, 1998). These two goals are not always compatible. Attempts to resolve the tension between the needs of the majority of students who will not become scientists and the needs of the minority who will have led to an emphasis on teaching science through enquiry (National Academy of Science, 1995; National Research Council, 2000), and new curriculum models (Millar, 2006) that address the needs of both groups. The emphasis in these frameworks and their associated curricula lies not on producing individuals who will be "producers" of scientific knowledge, i.e. the future scientists; rather, it is on educating all young people to become informed, critical users of scientific knowledge.

To understand and engage in critical discussions about issues that involve science and technology requires three domain-specific competencies. The first is the ability to provide explanatory accounts of natural phenomena, technical artefacts and technologies and their implications for society. Such an ability requires a knowledge of the fundamental ideas of science and the questions that frame the practice and goals of science. The second is the knowledge and understanding of scientific enquiry to: identify questions that can be answered by scientific enquiry; identify whether appropriate procedures have been used; and propose ways in which such questions might be answered. The third is the competency to interpret and evaluate data and evidence scientifically and evaluate whether the conclusions are justified. Thus, scientific literacy in PISA 2015 and PISA-D is defined by the three competencies to:

- explain phenomena scientifically
- interpret data and evidence scientifically
- evaluate and design scientific enquiry.

All of these competencies require knowledge. Explaining scientific and technological phenomena, for instance, demands a knowledge of the content of science (hereafter, content knowledge). The second and third competencies, however, require more than a knowledge of what is known; they depend on an understanding of how scientific knowledge is established and the degree of confidence with which it is held. Some have argued for teaching what has variously been called "the nature of science" (Lederman, 2006), "ideas about science" (Millar and Osborne, 1998) or "scientific practices" (National Research Council, 2012). Recognising and identifying the features that characterise scientific enquiry require a knowledge of the standard procedures that underlie the diverse methods and practices used to establish scientific knowledge (hereafter, procedural knowledge). Finally, the competencies require epistemic knowledge – an understanding of the rationale for the common practices of scientific enquiry, the status of the knowledge claims that are generated, and the meaning of foundational terms, such as theory, hypothesis and data.

Both procedural and epistemic knowledge are necessary to identify questions that are amenable to scientific enquiry, to judge whether appropriate procedures have been used to ensure that the claims are justified, and to distinguish scientific issues from matters of values or economic considerations. This definition of scientific literacy assumes that, throughout their lives, individuals will need to acquire knowledge, not through scientific investigations, but through the use of resources such as libraries and the Internet. Procedural and epistemic knowledge are essential to decide whether the many claims of knowledge and understanding that pervade contemporary media are based on the use of appropriate procedures and are justified.

Box 4.1 Scientific knowledge: PISA 2015 terminology

This document is based upon a view of scientific knowledge as consisting of three distinguishable but related elements. The first of these and the most familiar is a knowledge of the facts, concepts, ideas and theories about the natural world that science has established. For instance, how plants synthesise complex molecules using light and carbon dioxide or the particulate nature of matter. This kind of knowledge is referred to as "**content knowledge**" or "knowledge of the content of science".

Knowledge of the procedures that scientists use to establish scientific knowledge is referred to as "**procedural knowledge**". This is a knowledge of the practices and concepts on which empirical enquiry is based such as repeating measurements to minimise error and reduce uncertainty, the control of variables, and standard procedures for representing and communicating data (Millar, Lubben, Gott, and Duggan, 1995). More recently these have been elaborated as a set of "concepts of evidence" (Gott, Duggan, and Roberts, 2008).

Furthermore, understanding science as a practice also requires "**epistemic knowledge**" which refers to an understanding of the role of specific constructs and defining features essential to the process of knowledge building in science (Duschl, 2007). Epistemic knowledge includes an understanding of the function that questions, observations, theories, hypotheses, models, and arguments play in science; a recognition of the variety of forms of scientific enquiry; and the role peer review plays in establishing knowledge that can be trusted.

A more detailed discussion of these three forms of knowledge is provided in the later section on scientific knowledge and in Figures 4.5, 4.6 and 4.7.

People need all three forms of scientific knowledge to perform the three competencies of scientific literacy. PISA 2015 and PISA-D focus on assessing the extent to which 15-year-olds are capable of displaying the three aforementioned competencies appropriately within in a range of personal, local/national (grouped in one category) and global contexts. (For the purposes of the PISA assessment, these competencies are only tested using the knowledge that 15-year-old students can reasonably be expected to have already acquired.) This perspective differs from that of many school science programmes that are dominated by content knowledge. Instead, the framework is based on a broader view of the kind of knowledge of science required of fully engaged citizens.

In addition, the competency-based perspective also recognises that there is an affective element to a student's display of these competencies: students' attitudes or disposition towards science will determine their level of interest, sustain their engagement, and may motivate them to take action (Schibeci, 1984). Thus, the scientifically literate person would typically have an interest in scientific topics; engage with science-related issues; have a concern for issues of technology, resources and the environment; and reflect on the importance of science from a personal and social perspective. This requirement does not mean that such individuals are necessarily disposed towards becoming scientists themselves rather, such individuals recognise that science, technology and research in this domain are an essential element of contemporary culture that frames much of our thinking.

These considerations led to the definition of scientific literacy used in PISA 2015 and PISA-D (see Box 4.2). The use of the term "scientific literacy", rather than "science", underscores the importance that the PISA science assessment places on the application of scientific knowledge in the context of real-life situations.

Box 4.2 The 2015 definition of scientific literacy

Scientific literacy is 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, which requires the competencies to:

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

The competencies required for scientific literacy

Competency 1: Explain phenomena scientifically

The cultural achievement of science has been to develop a set of explanatory theories that have transformed our understanding of the natural world (in this document, "natural world" refers to phenomena associated with any object or activity occurring in the living or the material world), such as the idea that day and night is caused by a rotating Earth, or the idea that diseases can be caused by invisible micro-organisms. Moreover, such knowledge has enabled us to develop technologies that support human life by, for example, preventing disease or enabling rapid human communication across the globe. The competency to explain scientific and technological phenomena is thus dependent on a knowledge of these major explanatory ideas of science.

Explaining scientific phenomena, however, requires more than the ability to recall and use theories, explanatory ideas, information and facts (content knowledge). Offering scientific explanations also requires an understanding of how such knowledge has been derived and the level of confidence we might hold about any scientific claims. For this competency, the individual requires a knowledge of the standard forms and procedures used in scientific enquiry to obtain such knowledge (procedural knowledge) and an understanding of their role and function in justifying the knowledge produced by science (epistemic knowledge).

Competency 2: Interpret data and evidence scientifically

Interpreting data is such a core activity of all scientists that some rudimentary understanding of the process is essential for scientific literacy. Initially, data interpretation begins with looking for patterns, constructing simple tables and graphical visualisations, such as pie charts, bar graphs, scatterplots or Venn diagrams. At a higher level, it requires the use of more complex data sets and the use of the analytical tools offered by spreadsheets and statistical packages. It would be wrong, however, to look at this competency as merely an ability to use these tools. A substantial body of knowledge is required to recognise what constitutes reliable and valid evidence and how to present data appropriately.

Scientists make choices about how to represent the data in graphs, charts or, increasingly, in complex simulations or 3D visualisations. Any relationships or patterns must then be read using a knowledge of standard patterns. Whether uncertainty has been minimised by standard statistical techniques must also be considered. All of this draws on a body of procedural knowledge. The scientifically literate individual can also be expected to understand that uncertainty is an inherent feature of all measurement, and that one criterion for expressing confidence in a finding is determining the probability that the finding might have occurred by chance.

It is not sufficient, however, to understand the procedures that have been applied to obtain any data set. The scientifically literate individual needs to be able to judge whether they are appropriate and the ensuing claims are justified (epistemic knowledge). For instance, many sets of data can be interpreted in multiple ways. Argumentation and critique are essential to determining which is the most appropriate conclusion.

Whether it is new theories, novel ways of collecting data, or fresh interpretations of old data, argumentation is the means that scientists and technologists use to make their case for new ideas. Disagreement among scientists is normal, not extraordinary. Determining which interpretation is the best requires a knowledge of science (content knowledge). Consensus on key scientific ideas and concepts has been achieved through this process of critique and argumentation (Longino, 1990). Indeed, it is a critical and sceptical disposition towards all empirical evidence that many would see as the hallmark of the professional scientist. The scientifically literate individual understands the function and purpose of argument and critique and why they are essential to the construction of knowledge in science. In addition, they should be able both to construct claims that are justified by data and to identify any flaws in the arguments of others.

Competency 3: Evaluate and design scientific enquiry

Scientific literacy implies that students have some understanding of the goal of scientific enquiry, which is to generate reliable knowledge about the natural world (Ziman, 1979). Data collected and obtained by observation and experiment, either in the laboratory or in the field, lead to the development of models and explanatory hypotheses that enable predictions that can then be tested experimentally. New ideas, however, commonly build on previous knowledge. Scientists themselves rarely work in isolation; they are members of research groups or teams that engage, nationally and internationally, in extensive collaboration with colleagues. New knowledge claims are always perceived to be provisional and may lack justification when subjected to critical peer review – the mechanism through which the scientific community ensures the objectivity of scientific knowledge (Longino, 1990). Hence, scientists have a commitment to publish or report their findings and the methods used in obtaining their evidence. Doing so enables empirical studies, at least in principle, to be replicated and results confirmed or challenged. However, measurements can never be absolutely precise; they all contain a degree of error. Much of the work of the experimental scientist is thus devoted to resolving uncertainty by repeating measurements, collecting larger samples, building instruments that are more accurate, and using statistical techniques that assess the degree of confidence in any result.

In addition, science has well-established procedures that are the foundations of any experiment to establish cause and effect. The use of controls enables the scientist to claim that any change in a perceived outcome can be attributed to a change in one specific feature. Failure to use such techniques leads to results where effects are confounded and cannot be trusted. Likewise, double-blind trials enable scientists to claim that the results have not been influenced either by the subjects of the experiment, or by the experimenter themselves. Other scientists, such as taxonomists and ecologists, are engaged in the process of identifying underlying patterns and interactions in the natural world that warrant a search for an explanation. In other cases, such as evolution, plate tectonics or climate change, scientists examine a range of hypotheses and eliminate those that do not fit with the evidence.

Facility with this competency draws on content knowledge, a knowledge of the common procedures used in science (procedural knowledge), and the function of these procedures in justifying any claims advanced by science (epistemic knowledge). Procedural and epistemic knowledge serve two functions. First, such knowledge is required by individuals to appraise scientific investigations and decide whether they have followed appropriate procedures and whether the conclusions are justified. Second, individuals who have this knowledge should be able to propose, at least in broad terms, how a scientific question might be investigated appropriately.

The evolution of the definition of scientific literacy in PISA

In PISA 2000 and 2003, scientific literacy was defined as:

> "…the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity." (OECD, 2000, 2004)

In 2000 and 2003, the definition embedded knowledge *of* science and understandings *about* science within the one term "scientific knowledge". The 2006 definition separated and elaborated the term "scientific knowledge" by dividing it into two components: "knowledge *of* science" and "knowledge *about* science" (OECD, 2006). Both definitions referred to the application of scientific knowledge to understanding, and making informed decisions about, the natural world. In PISA 2006, the definition was enhanced by the addition of knowledge of the relationship between science and technology – an aspect that was assumed but not elaborated in the 2003 definition.

> "For the purposes of PISA, scientific literacy refers to an individual's:
>
> - Scientific knowledge and use of that knowledge to identify questions, acquire new knowledge, explain scientific phenomena and draw evidence-based conclusions about science-related issues.
> - Understanding of the characteristic features of science as a form of human knowledge and enquiry.
> - Awareness of how science and technology shape our material, intellectual and cultural environments.
> - Willingness to engage in science-related issues, and with the ideas of science, as a reflective citizen." (OECD, 2006)

These ideas have evolved further in the PISA 2015 definition of scientific literacy. The major difference is that the notion of "knowledge *about* science" has been specified more clearly and split into two components – procedural knowledge and epistemic knowledge.

In 2006, the PISA framework was also expanded to include attitudinal aspects of students' responses to scientific and technological issues within the construct of scientific literacy. In 2006, attitudes were measured in two ways: through the student questionnaire and through items embedded in the student test. Discrepancies were found between the results from the embedded questions and those from the background questionnaire with respect to "interest in science" for all students and gender differences in these issues (OECD, 2009; see also: Drechsel, Carstensen and Prenzel, 2011). More important, embedded items extended the length of the test. Hence, in PISA 2015, attitudinal aspects are only measured through the student questionnaire; there are no embedded attitudinal items. PISA-D does not include the measurement of attitudes towards learning science.

As for the constructs measured within this domain, the first ("interest in science") and third ("environmental awareness") remain the same as in 2006. The second ("support for scientific enquiry") has been changed to a measure of "valuing scientific approaches to enquiry", which is essentially a change in terminology to better reflect what is measured.

In addition, the contexts in PISA 2015 have been changed from "personal, social and global" in the 2006 assessment to "personal, local/national and global" to make the headings more coherent.

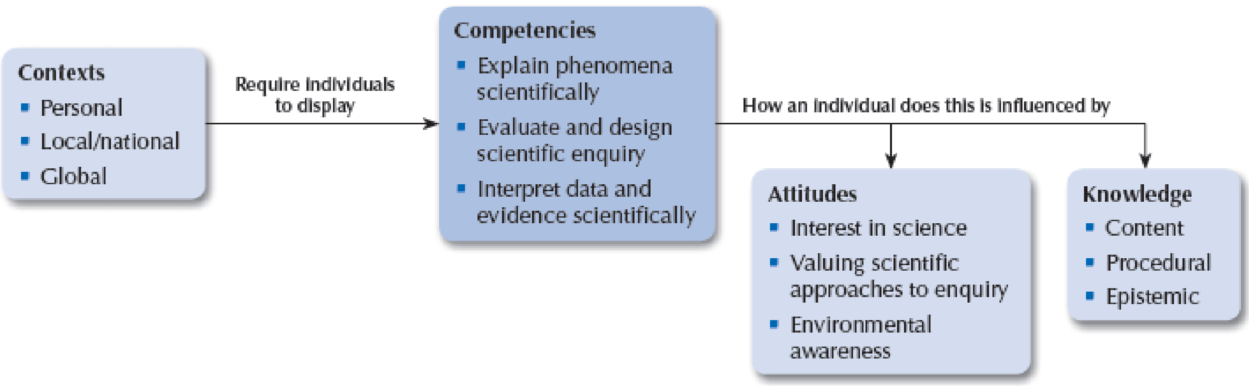
Organising the domain of science

The PISA 2015 definition of scientific literacy which is used in PISA-D consists of four inter-related aspects (see Figures 4.1 and 4.2).

Figure 4.1 Aspects of the scientific literacy assessment framework for PISA 2015

Contexts	Personal, local/national and global issues, both current and historical, which demand some understanding of science and technology.
Knowledge	An understanding of the major facts, concepts and explanatory theories that form the basis of scientific knowledge. Such knowledge includes knowledge of both the natural world and 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 justification for their use (epistemic knowledge).
Competencies	The ability to explain phenomena scientifically, interpret data and evidence scientifically, and .evaluate and design scientific enquiry.
Attitudes	A set of attitudes towards science indicated by an interest in science and technology, valuing scientific approaches to enquiry where appropriate, and a perception and awareness of environmental issues.

Figure 4.2 Inter-relations between the four aspects



Contexts of assessment items

PISA 2015 and PISA-D assesses scientific knowledge in contexts that are relevant to the science curricula of participating countries. Such contexts are not, however, restricted to the common aspects of participants' national curricula. Rather, the assessment requires evidence of the successful use of the three competencies required for scientific literacy in situations set in personal, local/national and global contexts.

Assessment items are not limited to school science contexts. In the PISA 2015 and PISA-D scientific literacy assessment, the items focus on situations relating to the self, family and peer groups (personal), to the community (local and national), and to life across the world (global). Technology-based topics may be used as a common context. Some topics may be set in historical contexts, which are used to assess students' understanding of the processes and practices involved in advancing scientific knowledge.

Figure 4.3 shows how science and technology issues are applied within personal, local/national and global settings. The contexts are chosen in light of their relevance to students' interests and lives. The areas of application are: health and disease, natural resources, environmental quality, hazards, and the frontiers of science and technology. They are the areas in which scientific literacy has particular value for individuals and communities in enhancing and sustaining quality of life, and in developing public policy.

Figure 4.3 Contexts in the PISA 2015 and PISA-D scientific literacy assessment

	Personal	**Local/National**	**Global**
Health and disease	Maintenance of health, accidents, nutrition	Control of disease, social transmission, food choices, community health	Epidemics, spread of infectious diseases
Natural resources	Personal consumption of materials and energy	Maintenance of human populations, quality of life, security, production and distribution of food, energy supply	Renewable and non-renewable natural systems, population growth, sustainable use of species
Environmental quality	Environmentally friendly actions, use and disposal of materials and devices	Population distribution, disposal of waste, environmental impact	Biodiversity, ecological sustainability, control of pollution, production and loss of soil/biomass
Hazards	Risk assessments of lifestyle choices	Rapid changes (e.g. earthquakes, severe weather), slow and progressive changes (e.g. coastal erosion, sedimentation), risk assessment	Climate change, impact of modern communication
Frontiers of science and technology	Scientific aspects of hobbies, personal technology, music and sporting activities	New materials, devices and processes, genetic modifications, health technology, transport	Extinction of species, exploration of space, origin and structure of the universe

The PISA science assessment is *not* an assessment of contexts. Rather, it assesses competencies and knowledge *in* specific contexts. These contexts are chosen on the basis of the knowledge and understanding that students are likely to have acquired by the age of 15.

Sensitivity to linguistic and cultural differences is a priority in item development and selection, not only for the sake of the validity of the assessment, but also to respect these differences among participating countries.

Scientific competencies

Figures 4.4a, 4.4b and 4.4c provide a detailed description of how students may display the three competencies required for scientific literacy. The set of scientific competencies in Figures 4.4a, 2.4b and 4.4c reflects a view that science is best seen as an ensemble of social and epistemic practices that are common across all sciences (National Research Council, 2012). Hence, all these competencies are framed as actions. They are written in this manner to convey the idea of what the scientifically literate person both understands and is capable of doing. Fluency with these practices is, in part, what 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 scientist, a scientifically literate student can be expected to appreciate the role and significance of these practices and try to use them.

Figure 4.4a PISA 2015 and PISA-D scientific competencies: Explain phenomena scientifically

Explain phenomena scientifically
Recognise, offer and evaluate explanations for a range of natural and technological phenomena demonstrating the ability to: • Recall and apply appropriate scientific knowledge. • Identify, use and generate explanatory models and representations. • Make and justify appropriate predictions. • Offer explanatory hypotheses. • Explain the potential implications of scientific knowledge for society.

Demonstrating the competency of *explaining phenomena scientifically* requires students to recall the appropriate content knowledge in a given situation and use it to interpret and explain the phenomenon of interest. Such knowledge can also be used to generate tentative explanatory hypotheses in contexts where there is a lack of knowledge or data. A scientifically literate person is expected to be able to draw on standard scientific models to construct simple representations to explain everyday phenomena, such as why antibiotics do not kill viruses, how a microwave oven works, or why gases are compressible but liquids are not, and use these to make predictions. This competency includes the ability to describe or interpret phenomena and predict possible changes. In addition, it may involve recognising or identifying appropriate descriptions, explanations and predictions.

For the purposes of assessing students who perform at Level 1, a more detailed description of this competency is defined beneath for PISA-D. All level 1 students should be able to demonstrate some ability to explain phenomena scientifically.

Explain phenomena scientifically for Level 1c
Recognise explanations for a limited range of the most simple natural and technological phenomena demonstrating the ability to: • Recall appropriate scientific knowledge
Explain phenomena scientifically for Level 1b
Recognise explanations for a range of simple or familiar natural and technological phenomena demonstrating the ability to: • Identify an explanatory model or representation • Recognise the potential implications of scientific knowledge for society and individuals
Explain phenomena scientifically for Level 1a
Recognise explanations for a range of simple or familiar natural and technological phenomena demonstrating the ability to: • Make appropriate predictions • Recognise an appropriate explanatory hypothesis • Recognize simple causal or correlational relationships

At Level 1c students can be required to:

- Identify what are the elements of a standard representation that are used in science. For instance, a question might present an unlabelled diagram of an object and students could be asked to add the appropriate labels from a list provided by the question.

- Recall appropriate scientific knowledge but not apply such knowledge. For instance, a student might be asked to identify which scientific phenomenon is being described in an item.

At Level 1b students can be required to:

- Recall appropriate scientific knowledge but not to apply such knowledge. For instance, a question might ask which one of several familiar scientific concepts from a list would explain a simple phenomenon described at the beginning of the question.

- Use a familiar piece of scientific knowledge. For instance, a question about the freezing point of water might ask students to determine whether water will freeze in a given context.

At Level 1a students can be required to:

- Make a simple prediction but not justify it. For instance, a question might ask which of several predictions might be correct, or students could be asked to predict the reading of an ammeter on a simple circuit where the reading on one ammeter is provided and the other is not.

- To identify from a list what the evidence is that supports a particular claim, e.g. that a rock is a sedimentary rock or that a whale is a mammal rather than a fish.

- Provide descriptive explanations of the properties of objects or substances – for instance that a rock must be sedimentary because it can be easily scratched.

The following requirements, however, would be considered too advanced and beyond the scope of a Level 1 competency; students would only be expected to attain partial credit on an item. Thus competency at this level would not require students to:

- Offer explanatory hypotheses or explain the potential implications of scientific knowledge for society

- Construct an explanation for why a given explanation might be flawed

- Offer explanatory hypotheses that require students to recall knowledge and draw an appropriate inference

- Provide a causal explanation for how a device works

- Identify an explanatory model in a question that requires the recall of more than two pieces of knowledge

- Provide explanations of unfamiliar phenomena

Figure 4.4b PISA 2015 and PISA-D scientific competencies: Interpret data and evidence scientifically

Interpret data and evidence scientifically
Analyse and evaluate scientific data, claims and arguments in a variety of representations and draw appropriate conclusions, demonstrating the ability to: • Transform data from one representation to another. • Analyse and interpret data and draw appropriate conclusions. • Identify the assumptions, evidence and reasoning in science-related texts. • Distinguish between arguments that are based on scientific evidence and theory and those based on other considerations. • Evaluate scientific arguments and evidence from different sources (e.g. newspapers, the Internet, journals).

A scientifically literate person should be able to interpret and make sense of basic forms of scientific data and evidence that are used to make claims and draw conclusions. Displaying this competency may require all three forms of scientific knowledge.

Those who possess this competency should be able to interpret the meaning of scientific evidence and its implications to a specified audience in their own words, using diagrams or other representations as appropriate. This competency requires the use of mathematical tools to analyse or summarise data, and the ability to use standard methods to transform data into different representations.

This competency also includes accessing scientific information and producing and evaluating arguments and conclusions based on scientific evidence (Kuhn, 2010; Osborne, 2010). It may also involve evaluating alternative conclusions using evidence; giving reasons for or against a given conclusion using procedural or epistemic knowledge; and identifying the assumptions made in reaching a conclusion. In short, the scientifically literate individual should be able to identify logical or flawed connections between evidence and conclusions.

The higher cognitive demand required to interpret data and evidence scientifically means that this competency is generally above Level 1c. More detailed descriptions of this competency for levels 1a and 1b are provided below for PISA-D.

Interpret data and evidence scientifically for Level 1b
Recognise a specific scientific claim, justification or data set in a simple or familiar context, demonstrating the ability to: • Identify the evidence, claim or justification in a science-related text • Identify simple patterns in data
Interpret data and evidence scientifically for Level 1a
Recognise specific scientific data, claims, and justifications in simple or familiar contexts and identify an appropriate conclusion demonstrating the ability to: • Recognise an appropriate conclusion that can be drawn from a simple set of data • Extract a specific piece of information from a scientific text • Identify a non-scientific argument • Interpret graphical and visual data • Identify simple causal or correlational relationships

At Level 1b students can be required to:

- Describe a simple trend in data, but not be asked to draw a conclusion based on the data. For instance, a question might be asked to identify how temperatures have changed over a period of time when provided data in a graph or table.
- Identify a claim or evidence or a reason in a science-related text. Alternatively, students could be asked to identify which is the claim, evidence or reasoning in a science text from a list that is provided.

At Level 1a students can be required to:

- State which one of several conclusions about a simple phenomenon that are drawn from a data set is the most appropriate using a deduction requiring one step.
- Given a simple table, graph or other form of data representation, identify which conclusion or prediction is correct, e.g. identifying trends in a graph where there is no extraneous information.
- Extract meaning from simple scientific texts, for instance, asking students to identify the states through which matter moves, e.g. about solids, liquids and gases.
- Identify whether the conclusion drawn from a table of results, graph or other form of data is justified or not, e.g. whether the interpretation drawn from a table of materials and the effect of a magnet on the material is correct.

However, the following requirements would be considered too advanced and beyond the scope of a Level 1 competency; students would only be expected to attain partial credit on an item. Thus competency at this level would not require students to:

- Distinguish between arguments that are based on scientific evidence or scientific theories and those that are based on other considerations.
- Evaluate two competing arguments from different sources (e.g. newspaper, Internet, journals).
- Analyse or interpret more than one data set in any question.
- Consider multiple pieces of evidence or multiple theories and whether the information supports one or more theories

Table 4.1 shows the desired distribution of items, by competency, for the PISA 2015 science assessment and for PISA-D. For science, the desired distributions for PISA-D are for the school-based instrument only, as science is not included in the out-of-school assessment.

Figure 4.4c PISA 2015 and PISA-D scientific competencies: Evaluate and design scientific enquiry

Evaluate and design scientific enquiry
Describe and appraise scientific investigations and propose ways of addressing questions scientifically demonstrating the ability to: • Identify the question explored in a given scientific study. • Distinguish questions that could be investigated scientifically. • Propose a way of exploring a given question scientifically. • Evaluate ways of exploring a given question scientifically. • Describe and evaluate how scientists ensure the reliability of data, and the objectivity and generalizability of explanations.

The competency of *evaluating and designing scientific enquiry* is required to evaluate reports of scientific findings and investigations critically. It relies on the ability to distinguish scientific questions from other forms of enquiry or recognise questions that could be investigated scientifically in a given context. This competency requires a knowledge of the key features of a scientific investigation – for example, what things should be measured, what variables should be changed or controlled, or what action should be taken so that accurate and precise data can be collected. It requires an ability to evaluate the quality of data, which, in turn, depends on recognising that data are not always completely accurate. It also requires the ability to determine whether an investigation is driven by an underlying theoretical premise or, alternatively, whether it seeks to determine patterns.

A scientifically literate person should also be able to recognise the significance of previous research when judging the value of any given scientific enquiry. Such knowledge is needed to situate the work and judge the importance of any possible outcomes. For example, knowing that the search for a malaria vaccine has been an ongoing programme of scientific research for several decades, and given the number of people who are killed by malarial infections, any findings that suggested a vaccine would be achievable would be of substantial significance.

Moreover, students need to understand the importance of developing a sceptical attitude towards all media reports in science. They need to recognise that all research builds on previous work, that the findings of any one study are always subject to uncertainty, and that the study may be biased by the sources of funding. This competency requires students to possess both procedural and epistemic knowledge but may also draw on their content knowledge of science, to varying degrees.

The higher cognitive demand required to evaluate and design scientific enquiry means that this competency is generally above Level 1c and attained only to a limited extent by Level 1b students. More detailed descriptions of this competency for levels 1a and 1b are provided below for PISA-D.

Evaluate and design scientific enquiry for Level 1b
Appraise simple scientific investigations, demonstrating the ability to: • Carry out a simple scientific procedure when provided explicit instructions • Determine which of several variables is the dependent variable in an investigation
Evaluate and design scientific enquiry for Level 1a
Appraise simple scientific investigations and recognise ways of addressing questions scientifically, demonstrating the ability to: • Identify the question explored in a simple scientific study • Distinguish a question that is possible to investigate scientifically from one that is not • Evaluate if one way of exploring a given question is scientifically appropriate • Recognise appropriate measures for a scientific quantity (units appropriate for measuring) • Identify a source of error in a measurement or a flaw in an experimental design

At Level 1b students can be required to:

- Determine which variables were changed, measured, or held constant when provided a description of a scientific investigation.

- Identify the appropriate instrument or units to measure a quantity from a selection of different instruments or units.

At Level 1a students can be required to:

- Identify the question that is being answered in a simple scientific investigation in which only one factor is varied at a time, for instance, by describing a study and then asking the student to explain what question is being answered.

- Identify which of several actions it might be best to undertake to answer a simple scientific question. For instance the question of "Where do woodlice live?" is best answered by pattern seeking, identification using criteria or fair testing.

- Propose specific measurements that might be needed to answer a simple scientific question. For instance, a question might ask which of several variables should be measured to investigate whether the length of a pendulum affects the time of swing. Alternatively a question might ask which of several variables should be controlled when conducting a simple investigation.

- From a list of several actions, identify which actions would reduce the error in an experiment. Such questions should be assessed using partial credit scoring.

- Identify a variable (dependent and independent variables and controlled variable) in a simple scientific enquiry that should be controlled or should be varied to answer a given question.

- Identify a simple flaw in an experimental design, e.g. a failure to control variables, taking a single measurement or measuring the wrong factor.

However, the following requirements would be considered too advanced and beyond the scope of a Level 1 competency; students would only be expected to attain partial credit on an item. Thus competency at this level would not require students to:

- Evaluate multiple ways of exploring a given question scientifically.
- Evaluate multiple ways that are proposed to ensure the reliability of data in an investigation.
- Explain why some data might be anomalous.
- Given a phenomenon, generate questions for investigation.

Table 4.1 Desired distribution of items, by competency

Competency	Percentage of items in PISA 2015	Percentage of items in PISA-D
Explain phenomena scientifically	40-50	40-50
Interpret data and evidence scientifically	30-40	30-40
Evaluate and design scientific enquiry	20-30	20-30

The desired distribution specifies the blueprint for the selection of items according to important aspects of the domain frameworks. Item selection is based on the assessment design as well as item characteristics related to a number of framework aspects, including competency, content, type of knowledge, and response formats, as well as consideration of the items' psychometric properties and appropriateness for this assessment. Following the assessment, the actual distributions of items across the framework aspects will be described in relation to the desired distributions. The extent to which the item pool for the assessment meets the framework specifications will be discussed in the technical report in the context of practical constraints in the item selection process.

Scientific knowledge

Content knowledge

Given that only a sample of the content domain of science can be assessed in the PISA 2015 and PISA-D scientific literacy assessment, clear criteria are used to guide the selection of the knowledge that is assessed. The criteria are applied to knowledge from the major fields of physics, chemistry, biology, earth and space sciences, and require that the knowledge:

- has relevance to real-life situations
- represents an important scientific concept or major explanatory theory that has enduring utility
- is appropriate to the developmental level of 15-year-olds.

It is thus assumed that students have some knowledge and understanding of the major explanatory ideas and theories of science, including an understanding of the history and scale of the universe, the particle model of matter, and the theory of evolution by natural selection. These examples of major explanatory ideas are provided for illustrative purposes; there has been no attempt to list comprehensively all the ideas and theories that might be considered fundamental for a scientifically literate individual.

Figure 4.5 shows the content knowledge categories and examples selected by applying these criteria. Such knowledge is required for understanding the natural world and for making sense of experiences in personal, local/national, and global contexts. The framework uses the term "systems" instead of "sciences" in the descriptors of content knowledge. The intention is to convey the idea that citizens have to understand concepts from the physical and life sciences, and earth and space sciences, and how they apply in contexts where the elements of knowledge are interdependent or interdisciplinary. Things viewed as subsystems at one scale may be viewed as whole systems at a smaller scale. For example, the circulatory system can be seen as an entity in itself or as a subsystem of the human body; a molecule can be studied as a stable configuration of atoms but also as a subsystem of a cell or a gas. Thus, applying scientific knowledge and exhibiting scientific competencies requires a determination of which system and which boundaries apply in any particular context.

Figure 4.5 Knowledge of the content of science

Physical systems that require knowledge of:
• Structure of matter (e.g. particle model, bonds) • Properties of matter (e.g. changes of state, thermal and electrical conductivity) • Chemical changes of matter (e.g. chemical reactions, energy transfer, acids/bases) • Motion and forces (e.g. velocity, friction) and action at a distance (e.g. magnetic, gravitational and electrostatic forces) • Energy and its transformation (e.g. conservation, dissipation, chemical reactions) • Interactions between energy and matter (e.g. light and radio waves, sound and seismic waves)
Living systems that require knowledge of:
• Cells (e.g. structures and function, DNA, plant and animal) • The concept of an organism (e.g. unicellular and multicellular) • Humans (e.g. health, nutrition, subsystems such as digestion, respiration, circulation, excretion, reproduction and their relationship) • Populations (e.g. species, evolution, biodiversity, genetic variation) • Ecosystems (e.g. food chains, matter and energy flow) • Biosphere (e.g. ecosystem services, sustainability)
Earth and space systems that require knowledge of:
• Structures of the Earth systems (e.g. lithosphere, atmosphere, hydrosphere) • Energy in the Earth systems (e.g. sources, global climate) • Change in Earth systems (e.g. plate tectonics, geochemical cycles, constructive and destructive forces) • Earth's history (e.g. fossils, origin and evolution) • Earth in space (e.g. gravity, solar systems, galaxies) • The history and scale of the universe and its history (e.g. light year, Big Bang theory)

Table 4.2 shows the desired distribution of items, by content of science, for PISA 2015 and PISA-D.

Table 4.2 Desired distribution of items, by content

System	**Percentage of items in PISA 2015**	**Percentage of items in PISA-D**
Physical	36	36
Living	36	36
Earth and space	28	28
Total	100	100

Procedural knowledge

A fundamental goal of science is to generate explanatory accounts of the material world. Tentative explanatory accounts are first developed and then tested through empirical enquiry. Empirical enquiry relies on certain well-established concepts, such as the notion of dependent and independent variables, the control of variables, types of measurement, forms of error, methods of minimising error, common patterns observed in data, and methods of presenting data.

It is this knowledge of the concepts and procedures that are essential for scientific enquiry that underpins the collection, analysis and interpretation of scientific data. Such ideas form a body of procedural knowledge that has also been called “concepts of evidence’ (Gott, Duggan and Roberts, 2008; Millar et al., 1995). One can think of procedural knowledge as knowledge of the standard procedures scientists use to obtain reliable and valid data. Such knowledge is needed both to undertake scientific enquiry and engage in critical reviews of the evidence that might be used to support particular claims. It is expected, for instance, that students will know that scientific knowledge has differing degrees of certainty associated with it, and so can explain why there is a difference between the confidence associated with measurements of the speed of light (which has been measured many times with ever more accurate instrumentation) and measurements of fish stocks in the North Atlantic or the mountain lion population in California. The examples listed in Figure 4.6 convey the general features of procedural knowledge that may be tested.

Figure 4.6. PISA 2015 and PISA-D procedural knowledge

Procedural knowledge
• The concept of variables, including dependent, independent and control variables; • Concepts of measurement e.g. quantitative (measurements), qualitative (observations), the use of a scale, categorical and continuous variables. • Ways of assessing and minimising uncertainty, such as repeating and averaging measurements. • Mechanisms to ensure the replicability (closeness of agreement between repeated measures of the same quantity) and accuracy of data (the closeness of agreement between a measured quantity and a true value of the measure). • Common ways of abstracting and representing data using tables, graphs and charts, and using them appropriately. • The control-of-variables strategy and its role in experimental design or the use of randomised controlled trials to avoid confounded findings and identify possible causal mechanisms. • The nature of an appropriate design for a given scientific question, e.g. experimental, field-based or pattern-seeking.

Epistemic knowledge

Epistemic knowledge refers to an understanding of the role of specific constructs and defining features essential to the process of knowledge building in science (Duschl, 2007). Those who have such knowledge can explain, with examples, the distinction between a scientific theory and a hypothesis or a scientific fact and an observation. They know that models, whether representational, abstract or mathematical, are a key feature of science, and that such models are like maps rather than accurate pictures of the material world. These students can recognise that any particle model of matter is an idealised representation of matter and can explain how the Bohr model is a limited model of what we know about the atom and its constituent parts. They recognise that the concept of a “theory” as used in science is not the same as the notion of a “theory” in everyday language, where it is used as a synonym for a “guess” or a "hunch". Procedural knowledge is required to explain what is meant by the control-of-variables strategy; epistemic knowledge is required to explain *why* the use of the control-of-variables strategy or the replication of measurements is central to establishing knowledge in science.

Scientifically literate individuals also understand that scientists draw on data to advance claims to knowledge, and that argument is a commonplace feature of science. In particular, they know that some arguments in science are hypothetico-deductive (e.g. Copernicus' argument for the heliocentric system), some are inductive (the conservation of energy), and some are an inference to the best explanation (Darwin's theory of evolution or Wegener's argument for moving continents). They also understand the role and significance of peer review as the mechanism that the scientific community has established for testing claims to new knowledge. As such, epistemic knowledge provides a rationale for the procedures and practices in which scientists engage, a knowledge of the structures and defining features that guide scientific enquiry, and the foundation for the basis of belief in the claims that science makes about the natural world.

Figure 4.7 represents what are considered to be the major features of epistemic knowledge necessary for scientific literacy.

Figure 4.7 PISA 2015 and PISA-D epistemic knowledge

Epistemic knowledge

The constructs and defining features of science. That is:

- The nature of scientific observations, facts, hypotheses, models and theories.
- The purpose and goals of science (to produce explanations of the natural world) as distinguished from technology (to produce an optimal solution to human need), and what constitutes a scientific or technological question and appropriate data.
- The values of science, e.g. a commitment to publication, objectivity and the elimination of bias.
- The nature of reasoning used in science, e.g. deductive, inductive, inference to the best explanation (abductive), analogical and model-based.

The role of these constructs and features in justifying the knowledge produced by science. That is:

- How scientific claims are supported by data and reasoning in science.
- The function of different forms of empirical enquiry in establishing knowledge, their goal (to test explanatory hypotheses or identify patterns) and their design (observation, controlled experiments, correlational studies).
- How measurement error affects the degree of confidence in scientific knowledge.
- The use and role of physical, system and abstract models and their limits.
- The role of collaboration and critique, and how peer review helps to establish confidence in scientific claims.
- The role of scientific knowledge, along with other forms of knowledge, in identifying and addressing societal and technological issues.

Epistemic knowledge is most likely to be tested pragmatically in a context where a student is required to interpret and answer a question that requires some of this type of knowledge rather than assessing directly whether they understand the features detailed in Figure 4.7. For example, students may be asked to identify whether the conclusions are justified by the data, or what piece of evidence best supports the hypothesis advanced in an item and explain why.

Table 4.3 describes the desired distribution of items by type of knowledge for PISA 2015 and PISA-D.

Table 4.3 Desired distribution of items, by type of knowledge

Knowledge	Percentage of items in PISA 2015	Percentage of items in PISA-D
Content	54-66	54-66
Procedural	19-31	19-31
Epistemic	10-22	10-22

The desired balance, by percentage of items, among the three knowledge components – content, procedural and epistemic – for PISA 2015 and PISA-D is shown in Table 4.4.[1]

Table 4.4 Desired distribution of items for knowledge

Knowledge types	Percentage of items in PISA 2015	Percentage of items in PISA-D
	Total over systems (physical, living, earth and space)	
Content	54-66	54-66
Procedural	19- 31	19- 31
Epistemic	10-22	10-22
Total over knowledge types	100	100

Assessing scientific literacy

Cognitive demand

A key new feature of the PISA 2015 framework that will also be used in PISA-D is the definition of levels of cognitive demand within the assessment of scientific literacy and across all three competencies of the framework. In assessment frameworks, item difficulty, which is empirically derived, is often confused with cognitive demand. Empirical item difficulty is estimated from the proportion of test-takers who solve the item correctly, and thus assesses the amount of knowledge held by the test-taker population, whereas cognitive demand refers to the type of mental processes required (Davis and Buckendahl, 2011). Care needs to be taken to ensure that the depth of knowledge required, i.e. the cognitive demand test items, is understood explicitly by the item developers and users of the PISA framework. For instance, an item can have high difficulty because the knowledge it is testing is not well known, but the cognitive demand is simply recall. Conversely, an item can be cognitively demanding because it requires the individual to relate and evaluate many items of knowledge – each of which is easily recalled. Thus, not only should the PISA test instrument discriminate in terms of performance between easier and harder test items, the test also needs to provide information on how students across the ability range can deal with problems at different levels of cognitive demand (Brookhart and Nitko, 2011).

The competencies are articulated using a range of terms defining cognitive demand through the use of verbs such as “recognise”, “interpret”, “analyse” and “evaluate”. However, in themselves these verbs do not necessarily indicate a hierarchical order of difficulty that is dependent on the level of knowledge required to answer any item. Various classifications of cognitive demand schemes have been developed and evaluated since Bloom's Taxonomy was first published (Bloom, 1956). These have been largely based on categorisations of knowledge types and associated cognitive processes that are used to describe educational objectives or assessment tasks.

Bloom’s revised Taxonomy (Anderson and Krathwohl, 2001) identifies four categories of knowledge – factual, conceptual, procedural and meta-cognitive. This categorisation considers these forms of knowledge to be hierarchical and distinct from the six categories of performance used in Bloom’s first taxonomy – remembering, understanding, applying, analysing, evaluating and creating. In Anderson and

Krathwohl's framework, these two dimensions are now seen to be independent of each other, allowing for lower levels of knowledge to be crossed with higher order skills, and vice versa.

A similar framework is offered by Marzano and Kendall's Taxonomy (2007), which also provides a two-dimensional framework based on the relationship between how mental processes are ordered and the type of knowledge required. The use of mental processes is seen as a consequence of a need to engage with a task with meta-cognitive strategies that define potential approaches to solving problems. The cognitive system then uses either retrieval, comprehension, analysis or knowledge utilisation. Marzano and Kendall divide the knowledge domain into three types of knowledge, information, mental procedures and psychomotor, compared to the four categories in Bloom's revised Taxonomy. Marzano and Kendall argue that their taxonomy is an improvement upon Bloom's Taxonomy because it offers a model of how humans actually think rather than simply an organising framework.

A different approach is offered by Ford and Wargo (2012), who offer a framework for scaffolding dialogue as a way of considering cognitive demand. Their framework uses four levels that build on each other: recall, explain, juxtapose and evaluate. Although this framework has not been specifically designed for assessment purposes, it has many similarities to the PISA 2015 definition of scientific literacy and the need to make more explicit references to such demands in the knowledge and competencies.

Another schema can be found in the framework based on Depth of Knowledge developed by Webb (1997) specifically to address the disparity between assessments and the expectations of student learning. For Webb, levels of depth can be determined by taking into account the complexity of both the content and the task required. His schema consists of four major categories: level 1 (recall), level 2 (using skills and/or conceptual knowledge), level 3 (strategic thinking) and level 4 (extended thinking). Each category is populated with a large number of verbs that can be used to describe cognitive processes. Some of these appear at more than one level. This framework offers a more holistic view of learning and assessment tasks and requires an analysis of both the content and cognitive process demanded by any task. Webb's Depth of Knowledge (DOK) approach is a simpler but more operational version of the SOLO Taxonomy (Biggs and Collis, 1982) which describes a continuum of student understanding through five distinct stages of pre-structural, unistructural, multistructural, relational and extended abstract understanding.

All the frameworks described briefly above have served to develop the knowledge and competencies in the PISA 2015 and PISA-D Frameworks. In drawing up such a framework, it is recognised that there are challenges in developing test items based on a cognitive hierarchy. The three main challenges are that:

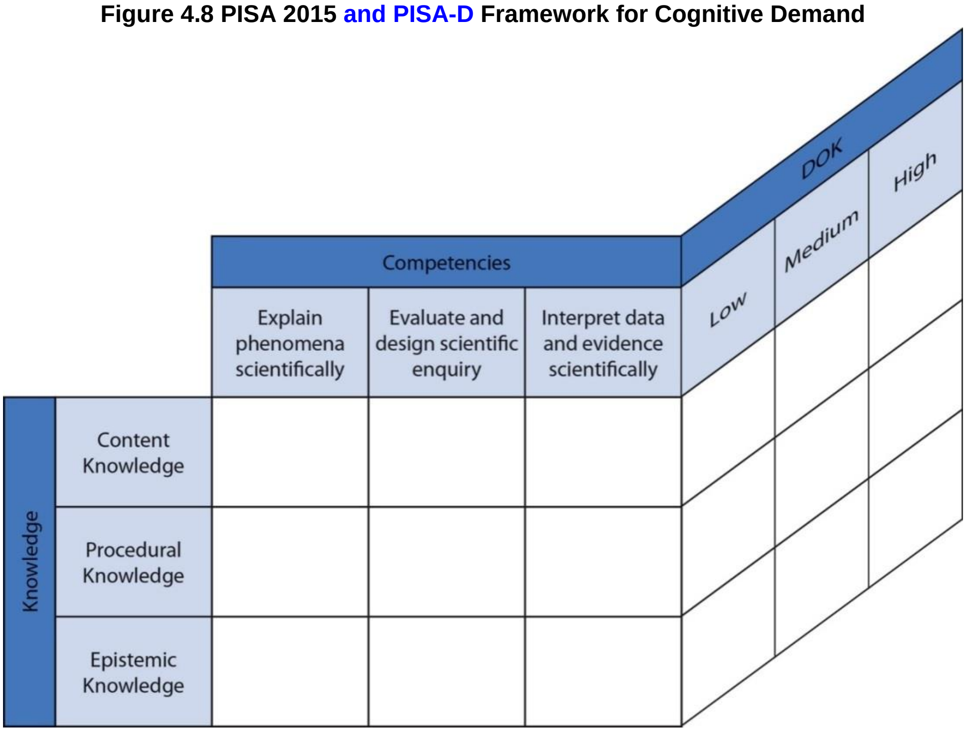
a) Too much effort is made to fit test items into particular cognitive frameworks, which can lead to poorly developed items.

b) Intended items (with frameworks defining rigorous, cognitively demanding goals) may differ from actual items (which may operationalise the standard in a much less cognitively demanding way).

c) Without a well-defined and understood cognitive framework, item writing and development often focuses on item difficulty and uses a limited range of cognitive processes and knowledge types, which are then only described and interpreted post hoc, rather than building from a theory of increasing competency.

The approach taken in this framework is to use an adapted version of Webb's Depth of Knowledge grid (Webb, 1997) alongside the desired knowledge and competencies. As the competencies are the central feature of the framework, the cognitive framework needs to assess and report on them across the student ability range. Webb's Depth of Knowledge Levels offer a taxonomy for cognitive demand that requires items to identify both the cognitive demand from the verbal cues that are used, e.g. analyse, arrange, compare, and the expectations of the depth of knowledge required.

Figure 4.8 PISA 2015 and PISA-D Framework for Cognitive Demand



The grid in Figure 4.8 provides a framework for mapping items against the two dimensions of knowledge and competencies. In addition, each item can also be mapped using a third dimension based on a depth-of-knowledge taxonomy. This provides a means of operationalising cognitive demand as each item can be categorised as making demands that are:

- **Low**
  Carry out a one-step procedure, for example recall a fact, term, principle or concept, or locate a single point of information from a graph or table.
- **Medium**
  Use and apply conceptual knowledge to describe or explain phenomena, select appropriate procedures involving two or more steps, organise/display data, interpret or use simple data sets or graphs.
- **High**
  Analyse complex information or data; synthesise or evaluate evidence; justify; reason, given various sources; develop a plan or sequence of steps to approach a problem.

Table 4.5. shows the real distribution of items by depth of knowledge for PISA 2015 (there was no desired distribution specified for the depth of knowledge categories). Since the PISA-D assessment design calls for a greater proportion of items measuring the lower end of the scale, this criterion will presumably affect the distribution of items across the three categories of depth of knowledge. Compared to the distribution of items in the PISA 2015 assessment, there will likely be a greater proportion of items classified as 'Low' or 'Medium' depth of knowledge than in the 'High' category in the PISA-D assessment.

Table 4.5 Distribution of items by depth of knowledge

Depth of knowledge	Percentage of items in PISA 2015	Percentage of items in PISA-D
Low	8	Not yet available
Medium	30	Not yet available
High	61	Not yet available
Total	100	100

Items that merely require recall of one piece of information make low cognitive demands, even if the knowledge itself might be quite complex. In contrast, items that require recall of more than one piece of knowledge, and require a comparison and evaluation of the competing merits of their relevance would be seen as having high cognitive demand. The difficulty of any item, therefore, is a combination both of the degree of complexity and range of knowledge it requires, and the cognitive operations that are required to process the item.

Therefore, the factors that determine the demand of items assessing science achievement include:

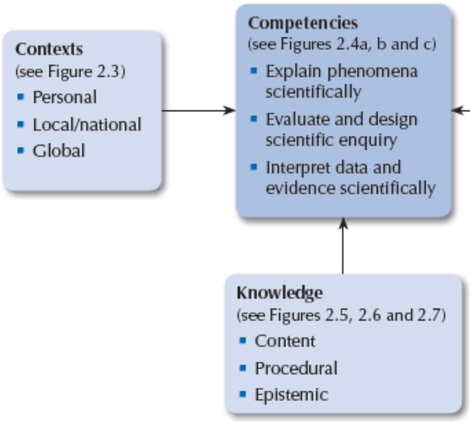
- The number and degree of complexity of elements of knowledge demanded by the item.
- The level of familiarity and prior knowledge that students may have of the content, procedural and epistemic knowledge involved.
- The cognitive operation required by the item, e.g. recall, analysis, evaluation.
- The extent to which forming a response is dependent on models or abstract scientific ideas.

This four-factor approach allows for a broader measure of scientific literacy across a wider range of student ability. Categorising the cognitive processes required for the competencies that form the basis of scientific literacy together with a consideration of the depth of knowledge required offers a model for assessing the level of demand of individual items. In addition, the relative simplicity of the approach offers a way to minimise the problems encountered in applying such frameworks. The use of this cognitive framework also facilitates the development of an a priori definition of the descriptive parameters of the reporting proficiency scales (see Figure 4.10).

Test characteristics

Figure 4.9 is a variation of Figure 4.2 that presents the basic components of the PISA framework for the 2015 scientific literacy assessment in a way that can be used to relate the framework with the structure and the content of assessment units. This may be used as a tool both to plan assessment exercises and to study the results of standard assessment exercises. As a starting point to construct assessment units, it shows the need to consider the contexts that will serve as stimulus material, the competencies required to respond to the questions or issues, the knowledge central to the exercise, and the cognitive demand.

Figure 4.9 A tool for constructing and analysing assessment units and items



A test unit is defined by specific stimulus material, which may be a brief written passage, or text accompanying a table, chart, graph or diagram. In units created for PISA 2015, the stimulus material may also include non-static stimulus material, such as animations and interactive simulations. The items are a set of independently scored questions of various types, as illustrated by the examples already discussed. Further examples can be found at the PISA website (www.oecd.org/pisa/).

PISA uses this unit structure to facilitate the use of contexts that are as realistic as possible, reflecting the complexity of real-life situations, while making efficient use of testing time. Using situations about which several questions can be posed, rather than asking separate questions about a larger number of different situations, reduces the overall time required for a student to become familiar with the material in each question. However, the need to make each score point independent of others within a unit needs to be taken into account. It is also necessary to recognise that, because this approach reduces the number of different assessment contexts, it is important to ensure that there is an adequate range of contexts so that bias due to the choice of contexts is minimised.

PISA 2015 test units require the use of all three scientific competencies and draw on all three forms of science knowledge. In most cases, each test unit assesses multiple competencies and knowledge categories. Individual items, however, assess only one form of knowledge and one competency.

The need for students to read texts in order to understand and answer written questions on scientific literacy raises an issue of the level of reading literacy that are required. Stimulus material and questions use language that is as clear, simple and brief, and as syntactically simplified as possible while still conveying the appropriate meaning. The number of concepts introduced per paragraph are limited. Questions within the domain of science that assess reading or mathematical literacy are avoided.

In PISA-D, for a better measurement of items at proficiency levels 1 and 2, items should only make the lower-level cognitive demands of recalling or recognizing appropriate knowledge, understanding the meaning of texts, applying that knowledge, and very simple data analysis drawing on either factual knowledge or foundational concepts (Anderson and Krathwohl, 2001; Webb 1997). In addition, whatever proficiency level they measure, items and language should wherever possible be simplified to reduce the cognitive load demanded of students (Sweller, 1994).

Response formats

Three classes of items will be used to assess the competencies and scientific knowledge identified in the framework. About one-third of the items will be in each of the three classes in PISA 2015 and PISA-D:

- 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 treated for scoring as a single item (the typical format in 2006)
  - selection of more than one response from a list
- constructed response: items calling for written or drawn responses.
  - Constructed-response items in scientific literacy 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 drawing (e.g. a graph or diagram). In a computer-based assessment, any such items is supported by simple drawing editors that are specific to the response required.

Reporting proficiency in science in PISA-D

To achieve the aims of PISA, scales must be developed to measure student proficiency. A descriptive scale of levels of competence needs to be based on a theory of how the competence develops, not just on a post-hoc interpretation of what items of increasing difficulty seem to be measuring. The 2015 draft framework therefore defined explicitly the parameters of increasing competence and progression, allowing item developers to design items representing this growth in ability (Kane, 2006; Mislevy and Haertel, 2006). Although comparability with the 2006 scale descriptors (OECD, 2007) has been maximised in order to enable trend analyses, the new elements of the 2015 framework, such as depth of knowledge, have also been incorporated. The scales have also been extended by the addition of a level "1b" to specifically address and provide a description of students at the lowest level of ability who demonstrate minimal scientific literacy and would previously not have been included in the reporting scales. The scales for the 2015 Framework therefore propose more detailed and more specific descriptors of the levels of scientific literacy, and not an entirely different model as shown in Figure 4.10. For PISA-D, the table describing the performance level expectations has been extended to include a new Level 1c.

The proposed level descriptors are based on the PISA 2015 Results Volume I (OECD, 2016) and offer a qualitative description of the differences between levels of performance. The factors used to determine the demand of items assessing science achievement that have been incorporated into this outline of the proficiency scales include:

- The number and degree of complexity of elements of knowledge demanded by the item
- The level of familiarity and prior knowledge that students may have of the content, procedural and epistemic knowledge involved
- The cognitive operation required by the item, e.g. recall, analysis, evaluation
- The extent to which forming a response is dependent on models or abstract scientific ideas

Figure 4.10 Summary description of the eight levels of science proficiency in PISA-D

Level	Lower score limit	Percentage of students across OECD countries at each proficiency level in PISA 2015	Percentage of students across 23 middle- and low-income countries at each proficiency level in PISA 2015	Descriptor
6	708	1.1%	0.1%	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	6.7%	0.8%	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	19.0%	5.0%	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	27.2%	15.5%	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	24.8%	28.3%	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 could be investigated scientifically.
1a	335	15.7%	31.5%	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	4.9%	15.7%	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.
1c	*	0.6% (percentage of students scoring below Level 1b in PISA 2015)	3.1% (percentage of students scoring below Level 1b in PISA 2015)	At Level 1c, students can use an element of basic or everyday scientific fact to identify a correct scientific explanation.

Note: * Will be available after the main study.

Items at the newly created Level 1c should be familiar to students' everyday lives or draw on ideas that permeate contemporary culture. All items should, whenever possible, attempt to draw on macroscopic phenomena that students may have experienced or observed or learned in the curriculum. Equally important is to have all items formulated in the simplest possible language. Sentences should be short and direct. Lengthy sentences, compound nouns, and complex phrasing should be avoided. Vocabulary used in the items must be carefully examined to avoid the use of academic language and, wherever possible, simplify the scientific language. Wherever possible, the cognitive processing should only require one-step reasoning and use simple data or descriptions.

In order to enter Level 1c performance a student must have the foundational skills required to:

- Read and comprehend simple sentences
- Use numeracy and basic computation
- Understand the basic components of tables and graphs
- Apply the basic procedures of scientific enquiry
- Interpret simple data sets

Testing scientific literacy in the out-of-school population

The scientific literacy domain is not included in the out-of-school PISA-D assessment due to practical reasons related with the instrument. On one hand, the total test allows a maximum of 50 minutes, which is not enough time to include an assessment of three domains, so it became necessary to choose only two. In deciding, it was taken into account that reading and mathematics literacy are considered as foundational skills and necessary for the development of scientific literacy skills. In addition, the target population was also considered. Science is the domain with the strongest link to school, so the least appropriate for a group that by definition has been exposed to less formal schooling. Thus, it was decided that reading and mathematics were the domains that should be included in the assessment for out-of-school 14 to 16 year-olds.

Examples of items for addressing the extended PISA-D science framework

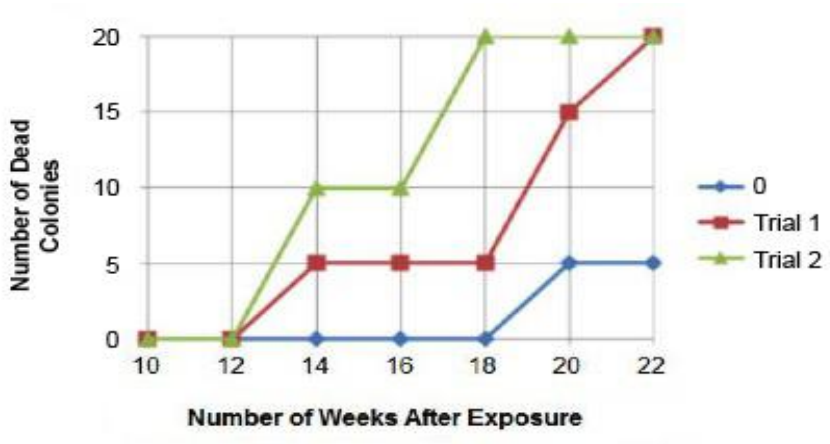
The following items illustrate the types of questions that can be asked of students at Level 1.

Sample item 1: Death of Bee Colonies – Level 1A

Scientists believe there are many causes why bee colonies die. One possibility is an insecticide that may cause bees to lose their sense of direction outside the hive.

Researchers tested whether this insecticide leads to the death of bee colonies. In a number of hives, they added the insecticide to the food of the bees for three weeks. All of the hives were given the same amount of food but the food had different amounts of insecticide in. Some hives were not given any insecticide.

None of the colonies died immediately. However, by week 14, some of the hives were empty. The following graph records the results:



What did the experiment test? Choose one of the responses below:

A. The experiment tested the effect of insecticide on the resistance of bees over time.
B. The experiment tested the effect of varying amounts of insecticide on the number of empty hives found over time.
C. The experiment tested the effect of the death of bee colonies on the resistance of bees to insecticide.
D. The experiment tested the effect of the death of bee colonies on the concentration of the insecticide.

Framework Categories	2015 Framework Extension
Competency	Evaluate and Design Scientific Enquiry
Full Description of Competency	Students must identify a question being asked in a simple scientific enquiry where only one factor is being varied at a time
Knowledge	Evaluate and Design Scientific Enquiry
Context	Local/National-Environmental Quality
Cognitive Demand	Low
Item format	Multiple Choice

Sample item 2: Fossil Fuels – Level 1A

Many power plants burn fuel that gives off carbon dioxide. Adding more carbon dioxide into the air has a negative impact on the climate. There are different strategies to reduce the amount of carbon dioxide added to the air.

One such strategy is to burn biofuels instead of fossil fuels.

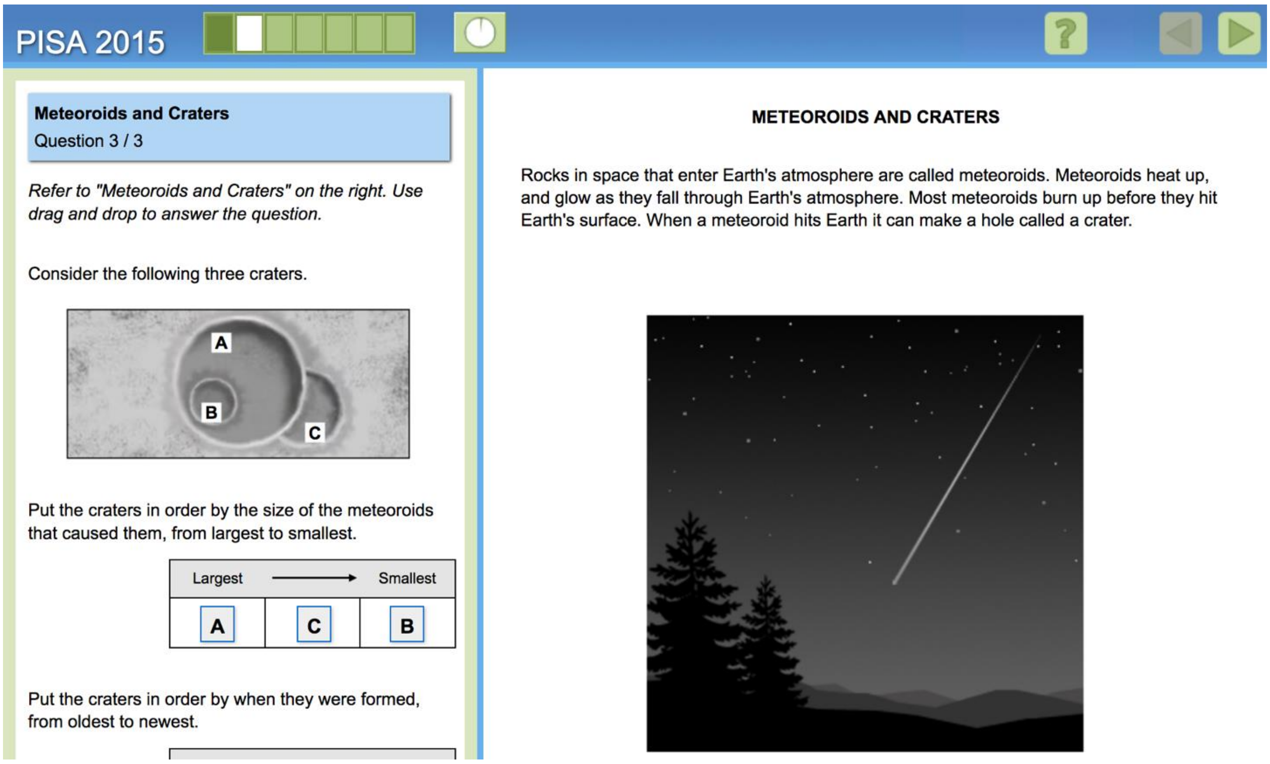
Another strategy involves trapping some of the carbon dioxide emitted by power plants and storing it deep underground or in the ocean. This strategy is called carbon capture.

Using biofuels does not have the same effect on levels of carbon dioxide in the air as using fossil fuels. Which of the statements below best explains why?

A. Biofuels do not release carbon dioxide when they burn.
B. Plants used for biofuels absorb carbon dioxide from the air as they grow.
C. As they burn, biofuels take in carbon dioxide from the air.
D. The carbon dioxide released by power plants using biofuels has different chemical properties than that released by power plants using fossil fuels.

Framework Categories	2015 Framework Extension
Competency	Explain Phenomena Scientifically
Full Description of Competency	Identify scientific concept appropriate to explain a phenomenon
Knowledge-System	Content-Physical
Context Global	Natural Resources
Cognitive Demand	Low
Item format	Multiple Choice

Sample Item 3: Meteoroids and Craters - Level 1b

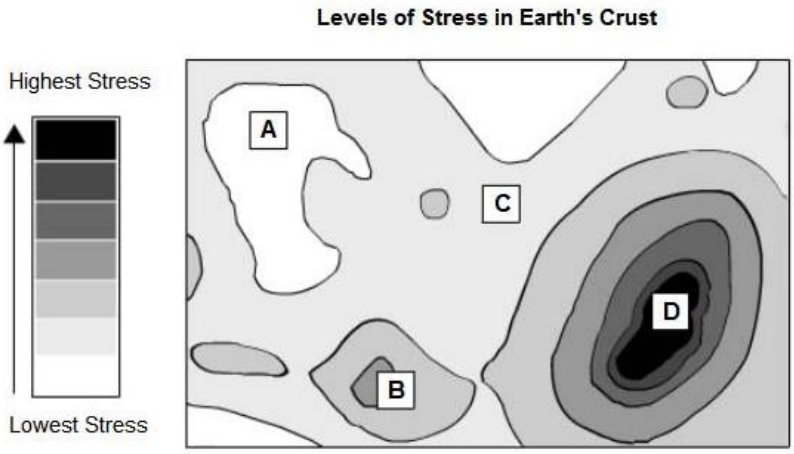


Framework Categories	2015 Framework Extension
Competency	Interpret Data and Evidence Scientifically
Full Description of Competency	Given a simple set of observations, draw a correct inference
Knowledge	Earth and Space
Context	Global
Cognitive Demand	Low
Item format	Multiple Choice

Sample item 4: Groundwater Extraction and Earthquakes-Level 1b

GROUNDWATER EXTRACTION AND EARTHQUAKES

Stress in Earth's Crust



The map above shows the levels of stress in Earth’s crust in a region. Four locations within the region are identified as A, B, C, and D. Each location is on or near a fault that runs through the region.

Which of the following correctly rank risk of earthquake from lowest to highest? Choose one of the answers below:

A. D, B, A, C
B. A, C, B, D
C. D, B, C, A
D. A, D, C, B

Framework Categories	2015 Framework Extension
Competency	Interpret Data and Evidence Scientifically
Full Description of Competency	Given a simple chart identify which conclusion is correct
Knowledge-System	Procedural
Context	Local/National-Hazards
Cognitive Demand	Low
Item format	Multiple Choice

NOTES

1. Because Science was the main domain in PISA 2015 this was reported separately for the three systems (physical, living, and earth and space). Since there are no subscales reported for PISA-D, the desired distribution for knowledge types is the total over all systems.

REFERENCES

American Association for the Advancement of Science. (1989). Science for all Americans: A Project 2061 report on literacy goals in science, mathematics and technology. Washington, DC: Author.

Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives. London, England: Longman.

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: W. H. Freeman and Company.

Biggs, J., & Collis, K. (1982). Evaluating the quality of learning: The SOLO taxonomy. New York, NY: Academic Press.

Bloom, B. S. (Ed.). (1956). Taxonomy of educational objectives: The classification of educational goals Handbook 1, Cognitive domain. London, England: Longmans.

Bøe, M. V., Henriksen, E. K., Lyons, T., & Schreiner, C. (2011). Participation in science and technology: Young people and achievement-related choices in late-modern societies. Studies in Science Education, 47(1), 37-72.

Bogner, F., & Wiseman, M. (1999). Toward measuring adolescent environmental perception. European Psychologist 4(3).

Brookhart, S. M., & Nitko, A. J. (2011). Strategies for constructing assessments of higher order thinking skills. In G. Schraw & D. R. Robinson (Eds.), Assessment of higher order thinking skills (pp. 327-359). Charlotte, NC: IAP.

Confederacion de Sociedades Cientificas de España. (2011). Informe ENCIENDE. Enseñanza de las Ciencias en la Didáctica Escolar para edades tempranas en España. Madrid, Spain: Author.

Davis, S. L., & Buckendahl, C. W. (2011). Incorporating cognitive demand in credentialing examinations. In G. Schraw & D. R. Robinson (Eds.), Assessment of higher order thinking skills (pp. 327-359). Charlotte, NC: IAP.

Drechsel, B., Carstensen, C., & Prenzel, M. (2011). The role of content and context in PISA interest scales – A study of the embedded interest items in the PISA 2006 science assessment. International Journal of Science Education, 33(1), 73-95.

Duschl, R. (2007). Science education in three-part harmony: Balancing conceptual, epistemic and social learning goals. Review of Research in Education, 32, 268-291.

Eagles, P. F. J., & Demare, R. (1999). Factors influencing children's environmental attitudes. Journal of Environmental Education, 30(4).

European Commission. (1995). White paper on education and training: Teaching and learning—towards the learning society. Luxembourg: Office for Official Publications in European Countries.

Fensham, P. (1985). Science for all: A reflective essay. Journal of Curriculum Studies, 17(4), 415-435.

Ford, M. J., & Wargo, B. M. (2012). Dialogic framing of scientific content for conceptual and epistemic understanding. Science Education, 96(3), 369-391.

Gardner, P. L. (1975). Attitudes to science. Studies in Science Education, 2, 1-41.

Gott, R., Duggan, S., & Roberts, R. (2008). Concepts of evidence. University of Durham. Downloaded from www.dur.ac.uk/rosalyn.roberts/Evidence/cofev.htm, Sept 23, 2012.

Kane, M. (2006). Validation. In R L. Brennan (Ed.), Educational measurement (4th ed.; pp. 17-64). Westport, CT: American Council on Education, Praeger Publishers.

Klopfer, L. E. (1976). A structure for the affective domain in relation to science education. Science Education, 60(3), 299-312.

Kuhn, D. (2010). Teaching and learning science as argument. [10.1002/sce.20395]. Science Education, 94(5), 810-824.

Lederman, N. G. (2006). Nature of science: Past, present and future. In S. Abell & N. G. Lederman (Eds.), Handbook of research on science education (pp. 831-879). Mahwah, NJ: Lawrence Erlbaum.

Longino, H. E. (1990). Science as social knowledge. Princeton, NJ: Princeton University Press.

Marzano, R. J., & Kendall, J. S. (2007). The new taxonomy of educational objectives. Thousand Oaks, CA: Corwin Press.

Millar, R. (2006). Twenty first century science: Insights from the design and implementation of a scientific literacy approach in school science. International Journal of Science Education, 28(13), 1499-1521.

Millar, R., Lubben, F., Gott, R., & Duggan, S. (1995). Investigating in the school science laboratory: Conceptual and procedural knowledge and their influence on performance. Research Papers in Education, 9(2), 207-248.

Millar, R., & Osborne, J. F. (Eds.). (1998). Beyond 2000: Science education for the future. London, England: King's College.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. Educational Measurement: Issues and Practice, 25(4), 6–20.

National Academy of Science. (1995). National science education standards. Washington, DC: National Academy Press.

National Research Council. (2000). Inquiry and the national science education standards. Washington DC: National Academy Press.

National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Washington, DC: Committee on a Conceptual Framework for New K-12 Science Education Standards, Board on Science Education, Division of Behavioral and Social Sciences and Education.

OECD. (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris, France: Author.

OECD. (2000). *Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical and scientific literacy*. Paris, France: Author.

OECD. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris, France: Author.

OECD. (2006). *The PISA 2006 assessment framework for science, reading and mathematics*. Paris, France: Author.

OECD. (2007). PISA 2006: *Science competencies for tomorrow's world: Volume 1: Analysis*. Paris, France: Author.

OECD. (2009). *PISA 2006 technical report*. Paris, France: Author.

OECD. (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264266490-en

Ormerod, M. B., & Duckworth, D. (1975). Pupils' attitudes to science. Slough, England: National Foundation for Education Research.

Osborne, J. F. (2010). Arguing to learn in science: The role of collaborative, critical discourse. Science, 328, 463-466.

Osborne, J. F., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. International Journal of Science Education, 25(9), 1049–1079.

Rickinson, M. (2001). Learners and learning in environmental education: A critical review of the evidence. Environmental Education Research, 7(3).

Rychen, D. S., & Salganik, L. H. (Eds.). (2003). Definition and selection of key competencies: Executive summary. Göttingen, Germany: Hogrefe.

Schibeci, R. A. (1984). Attitudes to Science: An update. Studies in Science Education, 11, 26-59.

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2005). Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss (Jahrgangsstufe 10).

Tai, R. H., Qi Liu, C., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. Science, 312, 1143-1145.

Taiwan Ministry of Education. (1999). Curriculum outlines for 'Nature science and living technology'. Taipei, Taiwan: Ministry of Education.

UNESCO (2003). UNESCO and the international decade of education for sustainable development (2005–2015). UNESCO International Science, Technology and Environmental Education Newsletter, Vol. XXVIII, no. 1–2. Paris, France: UNESCO.

UNESCO (2005). International implementation scheme for the UN decade of education for sustainable development. Paris, France: UNESCO.

United Nations Environment Programme. (2012). 21 issues for the 21st century: Result of the UNEP foresight process on emerging environmental issues. Nairobi, Kenya: Author.

Weaver, A. (2002), Determinants of environmental attitudes: A five-country comparison. International Journal of Sociology, 32(1).

Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education. Washington, DC: Council of Chief State School Officers and National Institute for Science Education Research Monograph.

Wiliam, D. (2010). What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. Review of Research in Education, 34, 254-284.

Ziman, J. (1979). Reliable knowledge. Cambridge, England: Cambridge University Press.

CHAPTER 5. PISA FOR DEVELOPMENT CONTEXTUAL QUESTIONNAIRES FRAMEWORK

This chapter describes the framework and core content for the PISA for Development (PISA-D) contextual questionnaires, for both the school-based assessment and the out-of-school assessment. The chapter presents the content and aims of the instruments for students who were in school and in grade 7 or higher at the time of the assessment, who were in school but in a grade lower than grade 7, and also for youth who were out of school. The chapter also describes the teacher and school questionnaires that are used for the school-based assessment and the instruments used for the out-of-school population: a questionnaire for the parents or the person most knowledgeable about the youth, and a household observation questionnaire.

As noted in Chapter 1, the focus of the PISA contextual questionnaires is on understanding how measures of student performance at age 15 are related to various aspects of school and classroom practice as well as to other related factors, such as economic, social and cultural context. The PISA-D questionnaires include these aspects and also cover a broader set of well-being outcomes and risk and protective factors, taking into account differences in life experiences of children in middle- and low-income countries, both of those who are in school and those who are not.

The PISA-D questionnaire framework uses the Educational Prosperity model (Willms, 2015) as an over-arching framework. It incorporates lessons from other international studies, inputs from the participating countries and many elements from the PISA questionnaires.

A review of the experience of middle- and low-income countries participating in PISA 2000 to 2015 shows that the PISA questionnaires do not always capture the most relevant contextual factors for these countries. For example, questions about school infrastructure and teaching and learning materials are related to student performance in high-income countries, but are often unrelated to differences in performance in middle-income countries (Lockheed, Bruer and Shadrova, 2015). In addition, the measure of economic, social and cultural status used by PISA does not adequately capture lower levels of parental education and income or the risk factors associated with poverty that are more frequent in low-income countries.

PISA-D enhances the contextual questionnaires to better measure factors that are more strongly related to student performance in middle- and low-income countries while maintaining comparability with PISA on a set of core indicators. For example, the questionnaires collect more detailed data on students' language of instruction at school, language at home and their socioeconomic status, as measured by home possessions and parents' education, literacy skills and participation in the labour force. The questionnaires also identify additional indicators of educational success beyond performance on the PISA test. These indicators are measured, for example, through questions about educational attainment, health and well-being, and student engagement in learning.

In addition to assessing student performance, PISA-D introduces an out-of-school assessment to collect data on youth that have not been eligible to sit the PISA school-based test. The out-of-school instruments gather much of the same data as the school-based instruments as well as data on barriers to school attendance and factors that may impede students' progress through school.

The PISA-D school-based instruments include student, teacher and school questionnaires. In contrast, PISA distributes questionnaires to students and schools and offers countries four optional questionnaires, including a computer familiarity questionnaire, an educational career questionnaire, a parent questionnaire and a teacher questionnaire. The PISA-D instruments for the out-of-school population include a youth interview, a questionnaire for their parents or the person most knowledgeable about the youth, and a household observation questionnaire that is completed by an interviewer.

This chapter presents the framework for the PISA-D contextual questionnaires. The first section defines the core of the PISA-D contextual assessment, explaining 1) the Education Prosperity framework that shaped the enhancements made to the contextual questionnaires for PISA-D, 2) the model for assessing equality and equity, 3) the approach to including the out-of-school population, 4) the potential of PISA-D to inform education policy, and 5) the selection and organisation of the core content of the PISA-D instruments. The second section of this chapter explores the full breadth of policy issues to be covered, structured in 14 modules and one complementary category, and explains how the modules have been implemented in PISA-D. The second section also includes a comparison of the policy issues covered by PISA-D and PISA, highlighting similarities and differences.

Defining the core of contextual assessment in PISA-D

The PISA-D framework is an adapted version of the Education Prosperity approach. It takes into account the goals of PISA-D, lessons from past PISA cycles and other international studies, recommendations from research literature and the priorities of the participating countries. This overarching framework maintains that policy makers in middle- and low-income countries need to be informed principally on the Prosperity Outcomes, Foundations for Success and student-level demographic factors for monitoring performance of their education system and assessing equality and equity of outcomes that are described in this document. In addition, the questionnaires include several teacher, school and system-level background measures that provide a context for the Prosperity Outcomes. The framework also proposes an approach for equality and equity for both the in-school and out-of-school populations. This framework is discussed below, specifying the constructs and measures and providing arguments that support the choice of core content for PISA-D.

Educational Prosperity

The PISA-D questionnaire framework draws on the Educational Prosperity model (Willms, 2015) that follows a life-course approach to assessing children's outcomes, considering the effects of several factors over a student or youth's lifetime. The capacity of a society to develop young peoples' literacy skills and well-being depends on its ability to provide the right kinds of human and material resources to support healthy development from conception to childhood and beyond. Educational Prosperity refers to the success of the education system in developing children's cognitive skills and their social, emotional, physical and spiritual well-being. The term "prosperity" simply refers to the condition of experiencing success or thriving (Willms, 2015).

Educational Prosperity, as applied in PISA-D, considers development from conception to adolescence as the result of individuals' personal characteristics, their actions, their culture and the contexts in which they live (Mayer, 2009). It identifies a set of key outcomes, called "Prosperity Outcomes", for each of six stages of development, and a set of family, institutional and community factors, called "Foundations for Success", which drive these outcomes. The stages, which are described in Annex B, are prenatal, early development (ages 0 to 2), pre-primary (ages 3 to 5), early primary (ages 6 to 9), late primary and lower secondary (ages 10 to 15), and upper secondary (ages 16 to 18). Both the school-based and out-of-school components of PISA-D focus on the Prosperity Outcomes and the Foundations for Success for the fifth stage of the Educational Prosperity framework, while the out-of-school component also collects some data on earlier stages.

The approach has three explicit links to national and local policy and practice. First, it allows countries to align data collection with explicit goals at all levels of the system, from the Minister and his or her staff to the front-line educators, students and parents. The challenge for countries is to maintain a focus on the alignment between data and policy goals. Second, the data collected with this approach has immediate implications for educational policies that involve decisions about the allocation of resources and its implications for equity. Countries will have reliable data on a wide set of prosperity outcomes as well as the foundation factors. With reliable data on differences across groups in outcomes and access to foundations countries will be able to determine whether poor and marginalised populations are given equal opportunities to succeed at school and beyond. Third, the data collected will enable countries to set targets consistent with the UN Education 2030 framework and monitor progress towards them. Many policy issues in middle- and low-income countries concern long-standing structural features of schools, such as the incidence of grade repetition or the choice of the language of instruction for minority groups. Making progress in reaching the Education 2030 goals will require confronting these issues on the basis of solid evidence on how these structural features of education systems relate to the achievement and well-being of students on average and of specific groups at risk. Comparable data from other countries facing similar policy changes can also facilitate the identification of policy options that can yield the desired results. With the Education Prosperity framework, PISA-D provides an infrastructure for analysing relationships between trends in outcomes and policy changes. The descriptive evidence from PISA usefully complements experimental policy evaluations and more qualitative assessments of the implementation of policy reforms.

Prosperity Outcomes

The framework for PISA-D conceptualises success as something cumulative, emphasising that development at age 15 is a product of children's environments and experiences since birth. The PISA contextual questionnaires framework emphasises understanding how measures of student performance, attitudes and beliefs at age 15 are related to various aspects of student background, and school and classroom practice. The PISA-D framework proposes a wider set of cognitive and non-cognitive outcomes and foundation factors to better measure the life experiences of in and out-of-school of children in middle- and low-income countries. The Prosperity Outcomes include measures of academic performance, educational attainment, engagement at school, and health and well-being.

The Educational Prosperity model was adapted to fit with the needs of the PISA-D participating countries, taking account of analysis of the results of middle- and low-income countries in PISA questionnaires, reviews of relevant international and regional studies and consultation with representatives of the participating countries.

The Educational Prosperity model distinguishes four processes that determine how success accumulates from one stage to the next: biological embedding, Foundations for Success, cumulative development and institutional selection. The Foundations for Success are described below, while a description of the other processes is available in Annex C, and a description of the effects associated with each of the four processes along the Educational Prosperity pathway is available in Annex D.

Foundations for Success

The Foundations for Success are factors that affect children's outcomes at each stage of development. For example, from age 2 to age 5, children's development is affected by parents' engagement with the child and intra-family relations as well as by the quality of care at home and in early childhood centres. They are considered to be universal in that they are necessary conditions for success at each stage of development. The selection of the foundation factors was based on theory and a large body of research that provides evidence of the effects of each factor on student outcomes.

Three additional criteria were considered in determining which factors to include as Foundations for Success: the factors must be potent, proximal, *and* pervasive. A "potent" factor is one that has a strong correlation with an outcome or set of outcomes. For example, the quality of classroom instruction is arguably the most important driver of student outcomes during the schooling period (Anderson, 2004; Rosenshine, 2010, Kyriakides, Christoforou and Charalambous, 2013; Creemers and Kyriakides, 2006).

A "proximal" factor is close to the outcome in the sense that its relationship with the outcome is not mediated through some other factor. For example, the quality of classroom instruction has a direct, positive relationship on student outcomes, without any intervening factors. "Principal leadership" is also an important factor and several studies have shown that it is correlated with student outcomes. However, it is not proximal because the "effects" of principal leadership are mediated through the school-related foundations factors, namely inclusive context, quality instruction, learning time and material resources. Thus, a jurisdiction may allocate resources to improving principal leadership, but this would only result in improved outcomes if it leads to improvements in quality instruction, increased learning time, and so on.

A “pervasive” factor is positively correlated with a wide range of outcomes, although the strength of the correlation may vary with each outcome. For example, the effects associated with an “inclusive school context” not only affect student’s academic performance, but also their educational attainment, their health and well-being, and their social, institutional and intellectual engagement.

Equality and equity

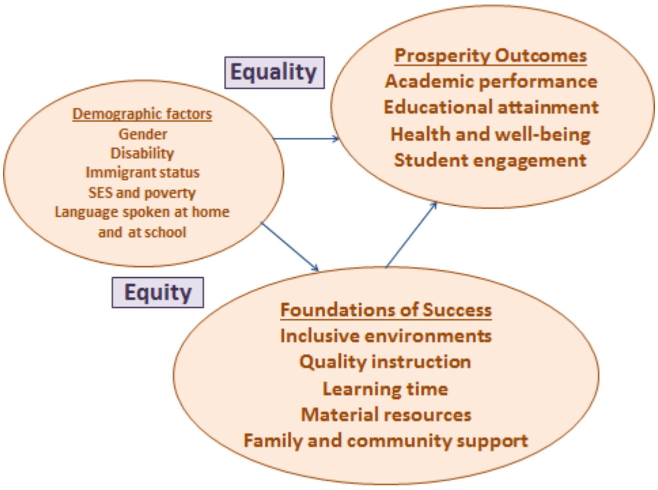
The terms “equality” and “equity” have been used by researchers and policy makers to denote several different concepts. These include, for example, the achievement gap between low- and high-status groups, differences in access to schooling, and the segregation of students into different types of schools and school programmes. Willms (2011) argued in the OECD’s 2011 *Education at a Glance* (OECD, 2011), that equality and equity should be defined as separate concepts and measured with a consistent approach.

PISA-D defines inequality as differences among sub-populations in the distribution of their educational outcomes, while the measure of equity, a normative concept, requires also an assessment of fairness based on the observed differences among sub-populations in their access to the resources and schooling processes that affect schooling outcomes. Equality is therefore measured by the differences among groups in the distribution of Prosperity Outcomes, which are performance, attainment, health and well-being and student engagement. Equity, on the other hand, also has to do with ensuring that all children benefit in the same way from school, and requires measures of whether children from different groups have fair access to the five foundations of success, which are inclusive environments, quality instruction, learning time, material resources and family and community support. Unfair access to the foundation factors increases inequalities in outcomes. For example, not providing the right level of support to disadvantaged children to attend schools regularly will inevitably result in socioeconomic inequalities in attainment and performance. Equity is a normative concept that is best assessed in relative terms – by comparing the levels of inequality in outcomes and in access to the foundations of success to those achieved by other countries, in comparable circumstances.

If we consider equality and equity in reading performance for students from differing socioeconomic backgrounds, for example, equality is assessed by examining the relationship between reading performance and socioeconomic status, while equity is assessed by also examining the relationship between socioeconomic status and the foundation factors that are considered core to learning how to read.

The PISA-D questionnaires collect information on several demographic factors that impact equality and equity and are relevant to both the in-school and the out-of-school populations. The framework focuses in particular on gender, disability, immigrant background, socioeconomic status and poverty and language. This model is characterised in Figure 5.1 (modified from Willms, Tramonte, Duarte and Bos, 2012).

Figure 5.1 A model for assessing equality and equity



When discussing equity it is important to always consider fairness and equality together with the need for quality. For example, a risk of policies focusing on equality without quality would be an education system where students from all social backgrounds have access to equally low quality education and perform equally poorly.

Assessing the out-of-school population

The first benchmark to assess equality and equity is whether all children are given the same opportunities to be in school and stay at school to acquire the skills they need for life. Some of the unique features of PISA-D are that it gathers information on how many youth are in school at ages 14 to 16 and the reasons why some youth have left school at that age. It also allows for the combination of data for the in-school and out-of-school populations.

PISA is aimed at 15-year-old students that are in grade 7 or above, leaving a large population of middle- and low-income countries out of the assessment. PISA-D incorporates these youth in the assessment through the out-of-school component which is conducted through a household survey. This component assesses 14-16-year-olds that are either not in school, or in school but in grade 6 or below, which represents approximately one-third of youth in the participating countries. Through its two components PISA-D includes students who are in school (those in grade 7 or above through the school-based component, those who are in grade 6 or below through the out-of-school component) and youth who are not in school. The youth that are out-of-school includes those that have never enrolled and those with some schooling, ranging from a few months to several years.

Box 5.1 Definition of access to schooling

The term "access" in education generally refers to whether schooling is freely available to children in a jurisdiction. The emphasis is on the *provision* of schooling, and it is incumbent upon governments and educational institutions to ensure that schools are available locally and that educational policies do not create barriers for attending school. In practical terms, however, "access" is gauged simply by measures of school attendance (e.g. UNESCO Institute for Statistics, 2006). This approach takes into account not only the supply of schooling, but also the cultural, social, religious, political and economic factors that affect the demand for schooling. In striving to improve school attendance, several governments have turned to demand-side initiatives, such as the provision of free meals, cash transfers to families which are conditional upon their child's attendance, and vouchers designed to increase school choice (Patrinos, 2007). Some definitions of "access" also incorporate the *quality* of school provision and in some cases are attached to a desired outcome. For example, the UN Sustainable Development Goal 4.1 states: "By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes" (United Nations, 2016). The statement calls not only for equal opportunities to attend school, but also equality of outcomes (relevant and effective learning outcomes) and equity of school provision (quality primary and secondary education).

The Educational Prosperity model and the approach taken in PISA-D identify two types of access: access as an outcome, which depends on both demand and supply and measured by attainment and learning outcomes; and access as a condition for success, which depends on supply and measured by the Foundations for Success related to the quality of schooling.

The PISA-D framework includes four key aspects of school quality: inclusive environments, quality instruction, learning time and material resources. A prerequisite to benefit from all school-related Foundations for Success is to be in school; therefore, access pertains to equity: do children from various sub-populations differ in their access to inclusive environments, quality instruction, learning time and material resources?

The PISA-D framework also includes academic performance and educational attainment, which refers to the extent to which children participate in school at various stages of their life-course and whether they are successful in making transitions from one stage of schooling to the next. Thus, access also has to do with equality: do children from various sub-populations differ in their distribution of attainment and performance?

PISA-D also includes measures pertaining to the barriers to attending school for out-of-school children, which can help discern the extent to which access is predominantly a supply- or demand-side issue for each country.

In general, out-of-school youth tend to be poorer than those attending school – many of them are in the lowest quintile. They are mainly from rural settings, and more likely to be girls. Youth with disabilities and those belonging to minority ethnic, linguistic or religious groups are also more likely to be out-of-school. All of these factors are usually confounded with poverty (Carr-Hill, 2015).

PISA-D's approach to measuring ESCS and poverty is especially important for the out-of-school population, as poverty is one of the main reasons, if not the main reason, for being out of school. As with the questionnaire for students attending school, the youth interview includes an extensive set of questions relevant to poverty and language spoken at home. The youth interview also includes several questions relevant to the demand for schooling, including questions about their work experience and support for their family as well as questions concerning perceived barriers to schooling.

The life-course approach assumes children's attainment is determined by various events and family circumstances that begin at conception and continue through to age 15. The questionnaire completed by the person most knowledgeable (PMK) about the youth asks about some elements of the early life-course foundations, such as the nutrition and health of the biological mother during pregnancy, and the engagement of the family during the preschool years. This provides data about the first four stages of development relevant to the accumulated effects of these factors on school attendance at ages 14 to 16.

Selecting and organising the core content

The instruments

The questionnaires for the school students include: a student questionnaire with 49 questions, a teacher questionnaire with 33 questions administered to the majority of teachers in each school, and a school questionnaire with 28 questions.

The field trial for the out-of-school youth include: an in-person interview with 102 questions, a questionnaire with 22 questions for the person most knowledgeable about the youth, and a household observation schedule completed by the interviewer. At the time of this publication the data for the out-of-school component main study was not yet available.

The distribution of questions across the elements of the Educational Prosperity model is shown in Figure 5.2, with red dots indicating questions that can be linked with PISA 2015 and blue dots indicating questions that are new to PISA-D.

Figure 5.2 Questions in the PISA-D questionnaires

PISA-D Field Trial - Number of Questions						
	Strand B	**Strand B**	**Strand B**	**Strand C**	**Strand C**	**Strand C**
	Student	**Teacher**	**School**	**Youth**	**Person Most Knowledgeable**	**Household**
Prosperity Outcomes						
Educational Attainment	🔴🔴🔴🔴			🔴🔴🔵🔵🔵🔵🔵🔵🔵🔵 🔵🔵🔵🔵🔵🔵🔵🔵🔵	🔵🔵🔵🔵	
Health and well-being	🔴🔵🔵🔵			🔴🔵🔵🔵		
Student engagement	🔴			🔵🔵	🔵	
Foundations for Success						
Inclusive environments	🔴🔵🔵🔵	🔵🔵	🔴🔴🔵🔵	🔵🔵🔵		
Quality instruction	🔴🔵🔵	🔵	🔵			
Learning time	🔴🔵🔵🔵🔵	🔵	🔴🔵	🔵🔵		
Material resources		🔵🔵🔵🔵🔵🔵	🔵🔵🔵🔵🔵			
Family and community support	🔵	🔵	🔵🔵	🔵	🔵	
Demographic factors to assess equity and equality						
Gender	🔴			🔴		
Language spoken at home	🔴🔵🔵🔵	🔵🔵🔵		🔴🔵🔵🔵🔵🔵		
Disability	🔵			🔵🔵		
Immigrant status	🔴			🔴🔴🔴🔴		
Socioeconomic status and poverty	🔴🔴🔴🔴🔴🔴🔴🔴 🔵🔵🔵🔵🔵🔵🔵🔵🔵🔵 🔵🔵		🔵🔵	🔴🔴🔴🔴🔴🔴🔴🔴🔴🔴 🔴🔵🔵🔵🔵🔵🔵🔵🔵🔵 🔵🔵🔵🔵🔵🔵🔵🔵🔵🔵 🔵🔵🔵🔵🔵🔵🔵🔵🔵🔵 🔵🔵🔵🔵🔵🔵🔵🔵🔵🔵 🔵🔵🔵🔵🔵🔵🔵🔵	🔵🔵🔵🔵🔵🔵🔵🔵🔵🔵	🔵🔵🔵🔵🔵🔵🔵🔵🔵🔵
Context factors		🔴🔴🔴🔴🔴🔴🔵🔵🔵🔵 🔵🔵🔵🔵🔵🔵🔵🔵🔵	🔴🔴🔴🔴🔴🔴🔴🔴 🔵🔵🔵🔵		🔴🔴🔴🔵🔵🔵	🔵🔵🔵🔵🔵🔵🔵

Total	**49**	**33**	**28**	**102**	**22**	**17**

Note: In the case of the out-of-school component, the instruments described correspond to the field trial because the main study instruments were not available at the time of this publication.

The questions for the PISA-D questionnaires were drawn mainly from PISA and complemented with questions from other international and regional studies or that were developed in consultation with the PISA-D participating countries. The criteria for selecting and developing items included the fit with the Educational Prosperity model, relevance as confirmed through analysis of the results of middle- and low-income countries in PISA questionnaires and reviews of relevant international and regional studies and consultation with representatives of the participating countries as well as maintaining links with PISA 2015.

All the items in the questionnaires were tested in a field trial. Questions were not retained for the main study if their psychometric properties (e.g. reliability, unidimensionality, completion of items and consistency across cultures) were inadequate. When there were two versions of a particular question, only one question was retained. To be retained for the main study questions also had to meet at least one of the following conditions:

1. Relevant to the measurement of ESCS common to PISA 2015 or new measures required to extend the scale to lower values of ESCS and to collect information on poverty;

2. Required for a measure of material resources;

3. Relevant for comparability with PISA 2015;

4. Required for coverage of all domains of the Educational Prosperity framework; or

5. Relevant to the classification of students into the five key sub-populations.

Core content for assessing Educational Prosperity

As noted above, the conceptual framework for the PISA-D questionnaires includes fourteen modules. These include four Prosperity Outcomes, five Foundations for Success, and five student demographic background factors relevant to assessing equality and equity (see Figure 5.2 above). It also includes a set of questions grouped under the category of context information that complement the variables included in the Educational Prosperity approach.

The content of these modules is discussed below.

Assessing Prosperity Outcomes

As noted above, Prosperity Outcomes include: academic performance, educational attainment, health and well-being and student engagement. These are each described briefly below.

Academic performance

The measures of academic performance in PISA-D are based on the assessments of performance in reading, mathematics and science. The frameworks for these assessments are described in Chapters 2, 3 and 4.

Educational attainment

Educational attainment – how far students go in school – is a key outcome for middle- and low-income countries that sits alongside measures of academic performance. Many of the key policy questions of middle- and low-income countries pertain to students' and families' demand for education, which depends on students' early learning experiences and their perceptions of its relevance, quality and long-term benefits. In many middle- and low-income countries, students do not attend school beyond the primary level.

A salient feature of middle- and low-income countries is that the distribution of 15-year-old students stretches well below grades 9 or 10. Another salient feature is that even though formally education might be technically compulsory, a large proportion of 15-year-old youth have dropped out. The primary aim in measuring attainment is to gain a better understanding of students' pathways to their current level of attainment and to also understand the reason for abandoning school, when this has happened.

Information about attainment is collected through all PISA-D questionnaires except for the teacher questionnaire and the household observations questionnaire, which is answered by the out-of-school youth interviewer. Like PISA, the PISA-D student and out-of-school youth questionnaires include questions about grade, early childhood education attendance and grade repetition; and the school questionnaire asks about grade retention policies and academic support services. PISA-D explores educational attainment in greater depth than PISA by asking students and out-of-school youth about long-term absenteeism and reasons for missing school for long periods. PISA-D further investigates the experience of out-of-school youth with questions about whether they work, their profession, hours worked per week, and wage or salary. Also unique to PISA-D, the parent questionnaire asks about their educational expectations for the out-of-school youth and factors that could hinder the youth's completion of compulsory education.[1]

The approach used by PISA-D to assess educational attainment is inspired by the framework set out by UIS and UNICEF (2005), which have been used to characterise the entire school-age population. In PISA-D they will be used to describe the levels of attainment of age 15 youth who are in school, and 14-16-year-olds who are out of school.

Data from students' current grade level, or in the case of out-of-school students the last grade completed, as well as data on students' birthdate and information on grade repetition, will be used to construct an ordinal variable describing five levels of attainment:

1. **On-track.** Students are in their expected grade, given their birthdate; that is, they started school on schedule and have not repeated a grade. In most cases, this would be grades 9 or 10.

2. **One year below expected grade.** These students have usually repeated a grade or were out of school for a prolonged period. They would typically be in grades 8 or 9.

3. **Two or three years below expected grade.** In most cases, these students have repeated two or three grades, but some may have started late or simply faded in and out of school for a year. They would typically be in grades 7 or 8.

4. **Enrolled in school but are four or five grades below the expected grade.** In most cases, these students will have repeated more than three times, but some may have started late or simply faded in and out of school for one year or more. They would typically be in grades 5 or 6.

5. **Not attending school, and their highest grade attained was five or more years below expected grade.** These students are not attending school, and their highest grade attained was

below grade 5 or A subcategory might have to be added to include children who are currently not attending school but who finalised grades 7, 8 or 9.

For those with attainment levels 2 through 5 the analyses will be extended to discern the stage of schooling when they fell off-track by one or more grades.

Health and well-being

The concept of well-being is very broad, and typically refers to the quality of people's life. Diener (2006) defines subjective well-being as "an umbrella term for the different valuations people make regarding their lives, the events happening to them, their bodies and minds, and the circumstances in which they live" (p. 400). PISA 2015 uses the following definition of well-being, that extends beyond students' subjective appraisal of their life quality: "Students' well-being refers to the psychological, cognitive, social and physical functioning and capabilities that students need to live a happy and fulfilling life" (OECD, 2017).

This module is based on the New South Wales Department of Education and Communities framework for student well-being (New South Wales Department of Education and Communities, 2015) that considers the following five domains: emotional, physical, social, cognitive, and spiritual well-being. The health and wellbeing module focuses on the first two of these domains, while social and cognitive wellbeing are included in other modules. Spiritual wellbeing is not included in the PISA-D framework.

Emotional well-being is the affective component of well-being – people's reactions to their experiences. It can be positive, such as people's overall rating of their happiness as used in the World Happiness Report (Helliwell, Layard, and Sachs, 2012), or negative, such as people's feelings of anxiety, depression, or fear. As in PISA, the PISA-D student and out-of-school youth questionnaires ask about general life satisfaction. PISA-D includes measures of anxiety and depression and questions about their physical and mental health during the past year.

Children's physical health is the key element of physical well-being. It is particularly important in middle- and low-income countries as children's health in these countries is more often compromised in ways that affect their educational outcomes due to hunger; physical and emotional abuse; chronic illnesses such as asthma, bronchitis, diabetes or epilepsy; and acute illnesses that cause children to miss school and fall behind (Carlson et al., 2008; MacLellan, Taylor and Wood, 2008). While PISA collects information on students' nutrition and physical activity, PISA-D asks about respondents' overall perception of their health and about their mental health during the past year.

Social well-being pertains to students' sense of belonging and their connectedness to others. In this framework, it is covered by elements of the student engagement and inclusive environment modules, primarily focusing on students' sense of belonging at school and their connectedness to others at school.

One of the elements of cognitive wellbeing is academic performance. In this framework this is covered as the main element of academic performance.

Box 5.2 Well-being in PISA 2015 and PISA-D

According to the Framework for the Analysis of Student Well-Being in the PISA 2015 Study (Borgonovi and Pál, 2016), the five dimensions of students' well-being captured in PISA 2015 are:

Cognitive well-being: The cognitive dimension of students' well-being refers to the skills and foundations students have to participate effectively in today's society, as lifelong learners, effective workers and engaged citizens. It comprises students' proficiency in academic subjects, their ability to collaborate with others to solve problems and their sense of mastery of school subjects. In PISA 2015 it is assessed as the level of subject-specific skills and competencies students have acquired, that are measured through the PISA tests, and their self-beliefs in those subjects, that are measured through the questionnaires.

In PISA-D this is assessed through the tests and included in the academic performance module, which is considered a Prosperity Outcome.

Psychological well-being. The psychological dimension of students' well-being includes students' evaluations and views about life, their engagement with school, and the goals and ambitions they have for their future. In PISA 2015 it includes students' self-reported psychological functioning, and covers life satisfaction – students' overall evaluation about life in general – and three aspects of education related to psychological functioning: 1) the goal setting and emotions related to *students' career and educational expectations*; 2) *achievement motivation* related to students' appreciation of the educational opportunities they have, an engagement with learning and an interest in acquiring knowledge; and 3) *test and learning anxiety*.

As in PISA, the PISA-D student and out-of-school youth questionnaires ask about general life satisfaction, and PISA-D includes a measure of emotional distress (severe anxiety and depression) and questions about their physical and mental health during the past year in the health and well-being module. Concerning education related elements, PISA-D explores educational attainment in greater depth than PISA by asking students and out-of-school youth about long-term absenteeism and reasons for missing school for long periods. PISA-D further investigates the experience of out-of-school youth with questions about whether they work, their profession, hours worked per week, and wage or salary. The out-of-school component also gathers information about youth's engagement in reading and writing literacy activities, such as how often they read a newspaper, magazine or book, write a text or email, etc.

Physical well-being: The physical dimension of students' well-being refers to students' health status, engagement in physical exercise and the adoption of healthy eating habits. PISA 2015 covers two aspects of students' lifestyles: physical activity and eating habits.

PISA-D focuses on physical health and assesses it as part of the module health and well-being.

Social well-being: The social dimension of students' well-being refers to the quality of their social lives including their relationship with their family, their peers and their teachers (positive or negative), and how they perceive their social life in school (positive or negative), and how they perceive their social life in school. PISA 2015 measures students' sense of belonging at schooland their relationships with their teachers, their peers and their parents.

PISA-D assesses social well-being within the student engagement and inclusive environment modules, primarily focusing on students' sense of belonging at school and their connectedness to others at school.

Material well-being: Material resources make it possible for families to better provide for their children's needs and for schools to support students' learning and healthy development. Households who live in poverty find it difficult to ensure that their children have access to the educational and cultural resources they need to thrive in school and to realise their potential. Material resources make it possible for families to better provide for their children's needs and for schools to support students' learning and healthy development. Households who live in poverty find it difficult to ensure that their children have access to the educational and cultural resources they need to thrive in school and to realise their potential. Research shows a strong link between material well-being in childhood and different dimensions of well-being in adulthood. Providing adequate resources to children is important not only because it is a pre-requisite for successful development but also because teenagers in poverty do not have adequate living and learning conditions to fulfil their personal goals. PISA contains a rich set of information on the types of resources students have at home and, most importantly, at school: human resources, material resources extracurricular activities.

PISA-D assesses material well-being through the socioeconomic status and poverty module and also through questions on material resources of schools. In both cases it creates new questions to better address the needs of middle-income countries.

The PISA 2015 questionnaire design does not attempt to clearly articulate and identify input and outcome indicators for the five well-being dimensions, so some dimensions focus on well-being inputs others on outcomes, without an integrated measurement approach. This is coherent with PISA-D classifying some of the PISA 2015 well-being factors as Prosperity Outcomes, others as Foundations for Success, and others as demographic factors related with equity and equality.

Student engagement

The PISA studies have examined students' interest and motivation in reading, mathematics and science, and their participation in activities related to the subject. For example, the OECD report, *Reading for Change: Performance and Engagement across Countries* examined students' motivation and interest in reading, and the time students spend reading for pleasure and reading diverse materials (OECD, 2002). PISA has also considered engagement more broadly, to refer to students' attitudes towards schooling and their participation in school activities (Willms, 2003).

Several studies have considered engagement to be a predictor of educational performance and attainment, and there is strong evidence that engagement is correlated with both performance and attainment (Willms, 2003). However, in PISA-D it is considered an important outcome in its own right, situated as a Prosperity Outcome alongside performance and attainment. A strong case can be made that the direction of causation is reversed (from performance to engagement) at certain stages of the school system. For example, children who do not make a successful transition from learning-to-read to reading-to-learn are more likely to become disaffected from school during the late primary and lower secondary years. Moreover, engagement is seen "as a *disposition* towards learning, working with others and functioning in a social institution" (Willms, 2003, p. 8), and as such is a key Prosperity Outcome that leads to life-long learning and the likelihood of becoming a productive member of society.

Like PISA, the PISA-D student and out-of-school youth questionnaires include a measure of institutional engagement, providing information on general attitudes towards school and learning outcomes as well as attitudes towards learning activities. Out-of-school youth are asked about student engagement based on their experience when attending school; and their parents are asked about their attitudes towards education as well.

The out-of-school component also gathers information about youth's engagement in reading and writing literacy activities, such as how often they read a newspaper, magazine or book, write a text or email, etc.

Assessing the Foundation of Success

As noted earlier in this chapter, the Foundations for Success are factors that affect children's outcomes at each stage of development. They are considered to be universal in that they are necessary conditions for success at each stage of development. The selection of the foundation factors for PISA-D was based on theory and a large body of research that provides evidence of the effects of each factor on student outcomes. The factors selected for PISA-D are: inclusive environments, quality instruction, learning time, material resources, and family and community support. These factors are each described briefly below. Some of the elements included in each factor are core to the factor, while other elements can be considered as supporting content.

Inclusive environments

Inclusive environments are classrooms, schools and broader communities that value and support inclusion. "Inclusion is a process of addressing and responding to the diversity of needs of all learners through increasing participation in learning, cultures and communities, and reducing exclusion within and from education. It involves changes and modifications in content, approaches, structures and strategies, with a common vision which covers all children of the appropriate age range and a conviction that it is the responsibility of the regular system to educate all children" (UNESCO, 2005, p. 13). UNESCO's (2009) policy guidelines provide a schema for measuring aspects of inclusion relevant to teachers' and principals' attitudes and values.

Inclusive environments are places in which all students can succeed. *All* means learners across the categorical boundaries of disability, social class, gender, ethnicity, national origin, sexual orientation and religion. *Succeed* means succeeding in terms of learning, as well as in terms of physical, social, emotional and spiritual outcomes (Willms, 2009). The provision of inclusive environments is a foundation for Educational Prosperity in middle- and low-income countries as it concerns the opportunities for children with disabilities; children from ethnic, linguistic and religious minorities; and girls to have equal access to schooling and a complete school experience, including opportunities to learn, engage in the social life of the school, and feel accepted by their peers and teachers.

Inclusive classroom and school practices affect students' sense of belonging at school, their participation in the social life of the school and their opportunities to learn. A reason for abandoning school is negative attitudes and responses to diversity and a failure to provide necessary accommodations to meet special learning needs. The metrics also needed to capture the attitudes and practices of teachers and principals. Inclusion requires teachers to be ambassadors for inclusion in their community by embracing and celebrating diversity, becoming skilled at meeting the needs of students with special needs and using new approaches to assessing learning (Riehl, 2000). At the system level, inclusion is concerned with the extent to which students from different sub-populations or ability are segregated into different schools or school programmes.

For the school-based component, PISA-D collects information on inclusion from students, teachers and school principals. For the out-of-school component, it collects this information from the youth questionnaire, asking youth to describe their experience when they attended school.

As in PISA, PISA-D asks students to report on their sense of belonging at school. PISA-D further explores school climate with questions to students about the safety of their school, whether they feel safe at school and whether they have been sexually harassed at school; out-of-school youth are asked these same questions based on their experiences when they attended school.

PISA-D asks teachers a set of questions about their attitudes and practices towards teaching students with low literacy levels. Both PISA and PISA-D ask school principals about school policies concerning how students are admitted to the school and grouped for instruction as well as about the diversity of the school. PISA-D also asks about their attitude towards grade retention.

Quality instruction

Quality of instruction is the most important driver of student performance, but arguably the most difficult to define and measure. Anderson (2004) defined effective teachers as "those who achieve the goals they set for themselves or which they have set for them by others (e.g. ministries of education, legislators and other government officials, school principals)." (p. 22). His model assumes that teachers are aware of, understand and actively pursue goals; that they teach with a purpose – to facilitate learning – material which they consider worthwhile; and that their goals are concerned directly or indirectly with student learning. This perspective, that effective teachers are goal-oriented, is evident in virtually all of the contemporary models of effective instruction (Coe, Aloisi, Higgins and Major, 2014).

The "delivery of the lesson" and "interacting with students" are at the centre of Anderson's (2004) conceptual framework of teacher effectiveness. Four other elements of his framework – standards, learning units, classroom climate and culture, and classroom organisation and management – have effects that are mediated through lesson delivery and teacher-student interactions. All six elements have direct effects on student learning and engagement. Teachers' characteristics, including their professional knowledge, expectations and leadership, and students' characteristics, including their aptitudes, prior knowledge, and attitudes and values, are positioned behind the six core elements of the framework. In the language of

Educational Prosperity presented above, they are distal factors that have their effects through the proximal core elements and thus are included as contextual factors and not as a foundation for success in the PISA-D framework. This module is assessed through questions to students, teachers and school principals and is not assessed for the out-of-school component. Similar to PISA, the PISA-D student questionnaire includes measures on student-teacher interactions and assesses the classroom learning climate. PISA-D adds new questions on lesson delivery to gather information on the structure of lessons and teaching practices in mathematics lessons. PISA-D adds questions about their practices for teaching less able students. PISA-D also asks school principals about teachers' behaviours that could negatively impact on classroom climate.

Learning time

Learning time in middle- and low-income countries differs from that of high-income countries in several ways. In many cases, children of middle- and low-income families start school at a later age, they miss many days of school during the primary school period, and they are more likely to repeat grades. Many children work in part-time jobs outside the home from an early age. Moreover, there appears to be considerable variation in class time devoted to the three core PISA subjects, and curriculum coverage is not as deep. How learning time is measured in main PISA has changed through the cycles.

The school-based component of PISA-D captures learning time in and out of school. Similar to PISA, the PISA-D student questionnaire asks about reasons for loss in learning time due to student truancy, though it extends this measure to collect information about other reasons for missing school, such as being sick or having to look after others. PISA-D also collects information on reasons for reduced teaching time and the time students take to travel from their home to school. PISA-D asks teachers about the reasons they are absent and school principals about their policies regarding teacher absenteeism. PISA-D also asks school principals about the reasons for and amount of instructional time lost during the last year, as does PISA. This module is not assessed for the out-of-school component.

Material resources

Studies based on the *Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación* (LLECE) data by Murillo and Román (2011) and Willms and Somers (2001) suggest that in middle- and low-income countries school resources have substantial effects, even after taking into account the socioeconomic characteristics of students.

PISA-D's school-based component used a schema set out by Murillo and Román (2011), which distinguishes between basic services, didactic facilities and didactic resources:

- Basic services at the school include factors such as potable water, sewage services, bathrooms, electricity and telephones.
- Didactic facilities refer to places other than the classroom, for teaching and learning. These include, for example, school libraries, gymnasiums, art and music rooms, science laboratories, computer rooms and sports fields.
- Didactic resources can include very basic materials, such as textbooks and blackboards, through to computers in the school, laptop computers for students and teachers, and quality books in the library.

Whereas PISA asks about principals' perceptions of schools resources (lack or inadequate physical infrastructure, educational material) and whether the school lacks resources and collects information on the availability of ICT resources and internet connectivity, PISA-D questions to school principals focus on the availability and condition of school infrastructure and facilities as well as the availability of textbooks. The questions in PISA-D also distinguish between the availability of school resources and teachers' use of school resources.

Family and community support

The nature and extent of family and community support differs among countries, not only because of cultural differences, but also due to the large number of children living in poverty in many of the partner countries. PISA questionnaires include questions pertaining to family's static cultural capital, about the relations between parents and children, and between parents and other parents.

In consultations with countries participating in PISA-D there was a demand for questions about community support. Small and Newman (2001) describe two over-arching connections between community and families that are relevant for developing measures of community support. One considers the socialisation of children, with neighbourhoods moulding children into certain behavioural patterns. The other pertains to the access of resources that support parents in raising their children. This could include, for example, literacy programmes, recreation facilities and programmes, or interventions to combat drug use and violence. In PISA there have been traditionally few questions about the communities or neighbourhoods of students, though PISA 2015 includes a question to school principals about whether they identify and integrate resources from the community to strengthen school programmes, family practices, and student learning and development.

For PISA-D, the community comprises local neighbourhoods nested within the school catchment area, which is defined with geographic boundaries or by the area from which the school draws its students

PISA-D asks both students and out-of-school youth about the types of communication they have with their parents or someone in their family, whereas PISA asks about the student-parent relationship in terms of emotional warmth, stimulation, etc.

PISA-D asks teachers about families' involvement at school and asks school principals about how parent and community members or organisations contribute to the school, whereas PISA gathers information about school policies for parental involvement. PISA-D also asks parents of out-of-school youth about the type of support they provided to the youth in their early years, while PISA countries that distribute the optional parent questionnaire ask parents of PISA students about the support they provided to their children at the beginning of primary education and at age 15. PISA also asks parents about their participation in school-related activities and whether there are factors that have prevented them from participating in the activities.

Student-level demographic factors for assessing equality and equity

PISA-D focuses on the following measures pertaining to students' and families' backgrounds that are particularly relevant for low- and middle-income countries: gender, disability, immigrant status, socioeconomic status and poverty, and the language spoken at home and the language of instruction at school. Though ethnicity is a variable related with being out-of-school, it was not included as one of the five demographic factors because it is embedded within poverty, immigrant status, language spoken at home and language of instruction. Similarly, living in a rural area is not included as a core demographic factor for assessing equality and equity, as living in a rural area versus a larger community is confounded with student-level demographic factors. The school-level questionnaire includes a variable pertaining to the size of the community, which can be used to determine rural status of the school. This question is also part of the questionnaire for person most knowledgeable in the out-of-school component and can be used to determine rural status at the youth level. Therefore, one can discern whether levels of performance associated with rural status and other community types are attributable to student-level demographic factors and various foundation factors, such as material resources or quality instruction.

For each category, a single dichotomous variable was constructed which can be used to provide summary indices of equality and equity. However, broader measures were also developed to assess equality and equity in more detail, such as an extension to the PISA measure of socioeconomic status. The measures are described below.

Gender

Like PISA, the PISA-D question on gender simply asks students and youth whether they are male or female. PISA-D does not capture data about gender identity or sexual orientation.

Disability

PISA-D is the first PISA study to include self-reported measures pertaining to disability. The questions follow contemporary approaches to disability, which emphasise the extent to which a disability limits people in doing certain activities in a particular environment. For example, students are asked about whether a disability limited their participation in school activities, while out-of-school youth are asked about whether they have a disability or medical condition that limits their daily activities. Out-of-school youth who report having a disability are also asked about the nature of the disability.

Immigrant status

The measure of immigrant status is based on a long-standing approach used in PISA which is based on questions of students and youth about where they and their parents were born.

Socioeconomic status and poverty

Socioeconomic status (SES) refers to the position of a person or family in a hierarchical social structure, based on their access to, or control over, wealth, prestige and power (Mueller and Parcel, 1981; Dutton and Levine, 1989). Numerous studies have shown that a person's position on an SES hierarchy is related to a wide range of outcomes pertaining to their physical, economic and social well-being. SES affects people's educational opportunities, their access to certain labour markets, their exposure to health and crime risks, and their lifetime earnings.

The literature on child development in middle- and low-income countries focuses mainly on the risk factors associated with poverty, especially during the prenatal period and the early years. These include, for example, poor nutrition during pregnancy, a lack of stimulation during the early years, and stressful living conditions.

The measure of SES in PISA, which is called the index of economic, social and cultural status (ESCS), does not adequately capture lower levels of education and lower levels of income and wealth for the majority of students in middle- and low-income countries. PISA-D will extend this measure to lower levels of SES, keeping the link with the PISA measure. PISA-D will also gather specific information on poverty, and explore the possibility of developing a separate measure of it relevant to Educational Prosperity.

The PISA-D field trial student and out-of-school youth questionnaires include a large number of questions pertaining to family socioeconomic status and poverty, including the long-standing questions used in PISA which assess the highest educational level of the parents, the highest occupational status of parents, and an index of home possessions, which has been adjusted to middle- and low-income countries.

The questionnaire and interview also include new questions designed to capture youth's experience of poverty, including questions about material possessions, parents' education and engagement in literacy activities, and more detailed information about their parents' occupation. Like PISA, information about school meals is collected through the PISA-D school questionnaire.

Poverty is expected to be a fundamental characteristic of the out-of-school population. Unique to PISA-D, parents provide information about the out-of-school youth's prenatal and early childhood living conditions; and the household observation questionnaire includes questions about the youth's housing and neighbourhood including questions that discern whether the housing is in a rural or urban setting.

Language spoken at home and language of instruction

In several middle- and low-income countries, the students' first language differs from the language of instruction at school. Also, in some countries, the language of instruction during the primary grades, when children are learning to read, differs from the official language of instruction after the third or fourth grade. A further issue, which is more difficult to capture with a survey, is that in some schools, the teachers use the students' native language, or some combination of the native language and the official language of instruction.

PISA asks students, "What language do you speak at home most of the time?" This construct is extended for PISA-D in both the school-based and out-of-school components to include questions about the language of instruction at school and the language they first learnt to read. In addition, teachers are asked about which language they use during their lessons, as well as which language they use when talking with students.

Context factors

For the in-school component the PISA-D school and teacher questionnaires also gather data on some teacher, school and system-level background variables that are expected to help explain student outcomes. These variables were not included as one of the previous modules because they are not directly linked to one of the modules, or are variables related to more than one modules, or because are not universal, necessary, potent, proximal and pervasive to be included as one of the five foundations for success. Some of the questions used to assess these variables come from PISA 2015 and others were created to fit the needs of middle- and low-income countries.

Like PISA, PISA-D asks teachers about their age and sex, qualification, employment status, educational background, years of experience and professional development activities. PISA-D gathers new information about whether the teacher teaches multi-grade classrooms, holds multiple teaching jobs or works other jobs in addition to teaching, and a number of factors relevant to their pre-service training, socioeconomic status, and health and well-being.

Like PISA, the PISA-D teacher questionnaire asks questions about school leadership at the school where they work and job satisfaction. PISA-D adds a question about their satisfaction with specific aspects of their job, such as benefits and pay. Also unique to PISA-D, the teacher questionnaire asks about the proportion of students in their class that lack the literacy and numeracy skills to learn the curriculum.

School principals in both PISA and PISA-D are asked numerous questions on resources and management, including type of school location community, number of students, average class size, school management and funding, as well as how many full- and part-time teachers work at their school. PISA-D adds questions on school location and nearby hazards.

As in PISA, the contextual information collected through the PISA-D questionnaires is complemented by system-level data on contextual variables in educational systems. The system-level questionnaire used in PISA was adapted for use by PISA-D countries, and both versions capture data on the structure of national programmes, national assessments and examinations, instructional time, teacher training and salaries, educational finance (including enrolment), national accounts and population data (UIS, 2016).

For the out-of-school component there is no data collected on teachers and school level background variables, however the system-level data can provide relevant contextual information. Some background variables about the person who answers the person most knowledgeable questionnaire is also collected.

NOTES

1. While PISA offers an optional parent questionnaire, it is distributed to the parents of students sitting the PISA test in school and focuses on the parents' perceptions of and involvement in their child's school, their support for learning at home and school choice; and it acquires information on basic characteristics of the early childhood education and care arrangements of PISA participants, and reasons for attending or not attending early childhood education and care.

REFERENCES

Anderson, L.W. (2004). *Increasing teacher effectiveness* (2nd ed). Paris: UNESCO International Institute for Educational Planning.

Barker, D.J.P. (1994). The fetal origins of adult disease. *Fetal and Maternal Medicine Review, 6*, 71–80.

Barker, D.J.P. and Sultan, H. Y. (1995). Fetal programming of human disease. In M. A. Hanson, J. A. D. Spencer and C. H. Rodeck (Ed.), *The Fetus and Neonate*, vol. 3, *Growth*, pp. 255-274. Cambridge: Cambridge University Press.

Barr, R. G., Beek, P.J. and Calinoiu, N. (1999). Challenges to nonlinear modelling of infant emotion regulation in real and developmental time. In G. J. P. Savelsbergh, H. van der Maas and P. van Geert (Eds.), *Non-linear developmental processes* (pp. 15-37). Amsterdam: Elsevier Science Publishers.

Barrington, B.L. and Hendricks, B. (1989). Differentiating characteristics of high school graduates, dropouts and nongraduates. *Journal of Educational Research*, *82*(6), 309-319.

Bertram, T., Pascel, C., Bokhari, S., Gasper, M. and Holtermann, S. (2002). *Early Excellence Centre Pilot Programme second evaluation report 2000-2001* (Research brief RB 361). London: DfES.

Borgonovi, F. and J. Pál (2016), *A Framework for the Analysis of Student Well-Being in the PISA 2015 Study: Being 15 In 2015*, OECD Education Working Papers, No. 140, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jlpszwghvvb-en

Boyce, W.T. and Kobor, M. S. (2015). Development and the epigenome: The 'synapse' of gene-environment interplay. *Developmental Science*, *18*(1), 1-23.

Boyce, W.T., Sokolowski, M. B. and Robinson, G. E. (2012). Toward a new biology of social adversity. *PNAS*, *109*(2), 17143-48.

Burchinal, M.R., Howes, C. and Kontos, S. (2002). Structural predictors of child-care quality in child-care homes. *Early Childhood Research Quarterly*, *17*, 87-105.

Burchinal, M.R., Peisner-Feinberg, E., Bryant, D. M. and Clifford, R. M. (2000). Children's social and cognitive development and child-care quality: Testing for differential associations related to poverty, gender or ethnicity. *Applied Developmental Science*, *4*, 149-165.

Burns, T., Köster, F., and Fuster, M. (2016), *Education Governance in Action: Learning from Case Studies*, OECD Publishing, Paris. Coe, R., Aloisi, C., Higgins, S. and Major, L. E. (2014). *What makes great teaching? Review of the underpinning research.* Durham University: Centre for Evaluation and Monitoring.

Carr-Hill, R. (2015), "PISA for development technical strand c: Incorporating out-of-school 15- year-olds in the assessment", *OECD Education Working Papers*, No. 120, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/5js0bsln9mg2-en

Coleman, M. and Vaughn, S. (2000). Reading interventions for students with emotional/behavioral disorders. *Behavioral Disorders, 25,* 93-104.

Crane, J. (1991). The epidemic theory of ghettos and neighbourhood effects on dropping out and teenage child bearing. *American Journal of Sociology, 96*(5), 1226-1259.

Creemers, B.P M. and Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17,* 347–366.

Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives*, *15,* 213-238.

Cynader, M. S. and Frost, B. J. (1999). Mechanisms of brain development: Neuronal sculpting by the physical and social environment. In D. Keating and C. Hertzman (Eds.), *Developmental health and the wealth of nations: Social, biological and educational dynamics.* (pp. 153-184). New York: The Guilford Press.

Diener, E. (200*6). Guidelines for National Indicators of Subjective Well-Being and Ill-Being. Applied Research in Quality of Life, 1, 151-157.*

Dutton, D.B. and Levine, S. (1989). Overview, methodological critique, and reformulation. In J. P. Bunker, D. S. Gomby and B. H. Kehrer (Eds.) *Pathways to health* (pp. 29-69). Menlo Park, CA: The Henry J. Kaiser Family Foundation.

Ensminger, M.E. and Slusarcick, A. L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort, *Sociology of Education, 65*(2), 95-113.

Fagan, J. and Pabon, E. (1990). Contributions of delinquency and substance use to school dropout among inner-city youth. *Youth and Society, 21*(3), 306-354.

Francis, D.J., Shaywitz, S.E., Stuebing, K.K., Shaywitz, B.A. and Fletcher, J.M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psycholo*gy, *88,* 3-17.

Gilbert, S., Barr, L., Clark, W., Blue, M. and Sunter, D. (1993). *Leaving School: Results from a National Survey Comparing School Leavers and High School Graduates 18 to 20 Years of Age*. Ottawa: Statistics Canada and Human Resources and Labour Canada, 81-575E.

Gunnar, M.R. (1998). Quality of early care and buffering of neuroendocrine stress reactions: Potential effects on the developing human brain. *Preventive Medicine: An International Journal Devoted to Practice and Theory, 27*(2), 208-211.

Hanushek, E.a. and Woessmann, L. (2015). *The knowledge capital of nations: Education and the economics of growth*. Cambridge and London: MIT Press.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York: Routledge.

Helliwell, J.F., Layard, R. and Sachs, J. (eds.) (2012), *World Happiness Report*, Earth Institute, New York, Columbia University.

Howes, C., Phillips, D. A. and Whitebook, M. (1992). Teacher characteristics and effective teaching in child care: Findings from the National Child Care Staffing Study. *Child and Youth Care Forum, 21*(6), 399-414.

Janosz, M., LeBlanc, M., Boulerice, B. and Tremblay, R. (1997). Disentangling the weight of school dropout predictors: A test on two longitudinal samples. *Journal of Youth and Adolescence, 26*(6), 733-762.

Kagan, S.L. and Zigler, E.F. (1987). *Early schooling: The national debate*. New Haven, CT: Yale University Press.

Kozminsky, E. and Kozminsky, L. (2001). How do general knowledge and reading strategies ability relate to reading comprehension of high school students at different education levels? *Journal of Research in Reading*, *24*(2), 187-204.

Kyriakides, L., Christoforou, C. and Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. Teaching and Teacher Education, *36*, 143–152.

Leseman, P.P.M. (2002). *Early childhood education and care for children from low-income or minority backgrounds.* Paris: Organization for Economic Co-operation and Development.

Liu, D., Diorio, J., Tannenbaum, B., Caldji, C., Francis, D., Freedman, A., Sharma, S., Pearson, D., Plotsky, P.M. and Meaney, M.J. (1997). Maternal care, hippocampal glucocorticoid receptors, and hypothalamic-pituitary-adrenal responses to stress. *Science, 277*, 1659-62.

Lockheed, M., T. Prokic-Bruer and A. Shadrova (2015), *The Experience of Middle-Income Countries Participating in PISA 2000-2015*, The World Bank, Washington, D.C./OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264246195-en

Mayer, K.U. (2009). New directions in life course research. *Annual Review of Sociology, 35*, 413-33.

McEwan, B.S. (1998). Protective and damaging effects of stress mediators. *New England Journal of Medicine*, *338,* 171-179.

McEwan, B.S. and Schmeck, H.M. (1994). *The Hostage Brain.* New York: The Rockefeller Unversity Press.

Morris, D., Bloodgood, J. and Perney, J. (2003). Kindergarten predictors of first and second-grade reading achievement. *The Elementary School Journal*, *104*(2), 93-109.

Mueller, C. W. and Parcel, T.L. (1981). "Measures of Socioeconomic Status: Alternatives and recommendations." *Child Development*, *52*, 13-30.

Murillo, F.J. and Román, M. (2011). School infrastructure and resources do matter: analysis of the incidence of school resources on the performance of Latin American students. *School Effectiveness and School Improvement*, *22*(1), 29-50.

Nelson, E.E. and Panksepp, J. (1998). Brain substrates of infant-mother attachment: Contributions of opioids, oxytocin, and norepinephrine. *Neuroscience and Biobehavioral Reviews, 22*, 437-452.

New South Wales Department of Education and Communities (2015). The wellbeing framework for schools. Sydney: Department of Education and Communities.

Nonoyama-Tarumi, Y., Hughes, K. and Willms, J.D. (2015). The role of family background and school resources on elementary school students' mathematics achievement. *Prospects, 45*(3), 305-24.

Nonoyama-Tarumi, Y. and Willms, J.D. (2010). The relative and absolute risks of disadvantaged family background and low levels of school resources on student literacy. *Economics of Education Review, 29*(2), 214-224.

O'Reilly, T. and McNamara, D. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. *American Educational Research Journal*, *44*(1), 161-196.

OECD (2002). *Reading for change: Performance and engagement across countries: Results from PISA 2000,* Paris: OECD.

OECD (2017), *PISA 2015 Results (Volume III): Students' Well-Being*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264273856-en

Patrinos, P.A. (2007). *Demand-side financing in education.* Paris: UNESCO International Institute for Educational Planning.

Peisner-Feinberg, E.S. and Burchinal, M.R. (1997). Relations between preschool children's childcare experiences and concurrent development: The cost, quality, and outcomes study. *Merrill-Palmer Quarterly, 43***,** 451–477.

Peisner-Feinberg, E.S., Burchinal, M.R., Clifford, R.M., Culkin, M.L., Howes, C., Kagan, S.L. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development, 72,* 1534-1553.

Ramey, C.T. and Ramey, S.L. (1998). Early intervention and early experience. *American Psychologist*, *53,* 109-120.

Riehl, C.J. (2000). The principal's role in creating inclusive schools for diverse students: A review of normative, empirical, and critical literature on the practices of educational administration. *Review of Educational Research, 70*(1), 55-81.

Rosenshine, B. (2010). *Principles of instruction.* International Academy of Education. Geneva: UNESCO International Bureau of Education.

Rumberger, R.W. (1983). Dropping out of high school: The influence of race, sex and family background. *American Educational Research Journal*, *20*(2), 199-220.

Schweinhart, L.J. and Weikart, D.P. (1998). Why curriculum matters in early childhood education. *Educational Leadership, 55,* 57-60.

Shepard, L., Kagan, S.L. and Wurtz, E. (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: National Education Goals Panel.

Small, M.L. and Newman, K. (2001). Urban poverty after the truly disadvantaged: The rediscovery of the family, the neighbourhood, and culture. *Annual Review of Sociology, 27*(1), 23-45.

Snow, C.E., Burns, M.S. and Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington DC: National Academy Press.

Tramonte, L. and Willms, J.D. (2010). Cultural capital and its effects on education outcomes. *Economics of Education Review, 29*(2), 200-213.

Tramonte, L., Willms, J.D. and Gauthier, A. (2013). Engagement and Guidance: The Effects of Maternal Parenting Practices on Children's Development. *Journal of Family Issues.* DOI: 10.1177/0192513x13489959.

UNESCO (2005) *Guidelines for Inclusion: Ensuring Access to Education for All.* Paris: UNESCO.

UNESCO (2009). *Policy guidelines on inclusion in education*. Paris: UNESCO.

UNESCO Institute for Statistics and UNICEF (2005). *Children Out of School: Measuring exclusion from primary education.* Montreal: UIS.

UNESCO Institute for Statistics (2006). *Global Education Digest: Comparing Education Statistics across the World*. Montreal, Quebec, UIS.

UNESCO Institute of Statistics (2016), *Making Education Count for Development: Data Collection and Availability in Six PISA for Development Countries*, PISA, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264255449-en

United Nations (2016). Sustainable development knowledge platform.

Wehlage, G. G. and Rutter, R.A. (1986). Dropping out: how much do schools contribute to the problem? *Teachers College Record, 87*(3), 374-92.

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

Willms, J.D. (2003). *Student engagement at school: A sense of belonging and participation*. Paris: Organization for Economic Cooperation and Development.

Willms, J.D. (2006). *Learning divides: Ten policy questions about the performance and equity of schools and schooling systems.* Montreal: UNESCO Institute for Statistics.

Willms, J.D. (2009, October). *Classroom Diversity and Inclusion: The Educational Advantage***.** Plenary presentation at the Return to Salamanca – Global Conference on Inclusive Education. Salamanca, Spain.

Willms, J.D. (2010). School composition and contextual effects on student outcomes. *Teachers College Record, 112*(4), 1008-1037.

Willms, J.D. (2011). *An analysis plan on educational equality and equity: Recommendations for the OECD Education at a Glance.* Paper prepared for the OECD NESLI INES Network for the collection and adjudication of system-level descriptive information on educational structures, policies and practices (NESLI). Fredericton: UNB-CRISP.

Willms, J.D. (2015). *Educational Prosperity*. Fredericton: The Learning Bar Inc.

Willms, J.D., Friesen, S. and Milton, P. (2009). *What did you do in school today? Transforming classrooms through social, academic, and intellectual engagement.* (First National Report) Toronto: Canadian Education Association.

Willms, J.D. and Somers, M.-A. (2001). Family, classroom and school effects on children's educational outcomes in Latin America. *International Journal of School Effectiveness and Improvement*, *12*(4), 409-445.

Willms, J.D. and Tramonte, L. (2015). *Towards the development of contextual questionnaires for PISA for Development.* Paris: OECD.

Willms, J.D., Tramonte, L., Duarte, J. and Bos, S. (2012). Assessing educational equality and equity with large-scale assessment data: Brazil as a case study. Washington: *Inter-American Development Bank.*

Zill, N. (1999). "Promoting educational equity and excellence in kindergarten", in R.C. Pianta and M. Cox (Eds.), *The Transition to kindergarten*, (pp. 67-105), Baltimore, MD: Brookes Publishing Co.

ANNEX A. PISA FOR DEVELOPMENT EXPERTS AND SUPPORT STAFF

PISA-D expert group (DEG)
Jean-François Rouet (France), Reading Literacy Chair
Zbigniew Marciniak (Poland), Mathematical Literacy Chair
Jonathan Osborne (USA), Scientific Literacy Chair
Takuya Baba, Country Expert (Japan)
Nascira Ramia, Country Expert (Ecuador)
Njekwa Shula Mamunye, Country Expert (Zambia)

Country representative experts
Thai Heng, Cambodia
Non Dara, Cambodia
Ngor Penglong, Cambodia
Héctor Fabian Arebalo Mosquera, Ecuador
Daniel Javier Wing Reyna, Ecuador
Elizabeth Liendro, Ecuador
Nascira Ramia, Ecuador
Georgina Afre, Guatemala
María José del Valle, Guatemala
Mamadou Bachir Diaham, Senegal
Moussa Fall, Senegal
Oumar Bâ, Senegal
Takuya Baba, Zambia
Njekwa Shula Mamunye, Zambia
Menny Sibukome Nachibinga, Zambia
Felidah Mwanza Kiwala, Zambia

Pearson support staff
John de Jong, Programme Director
Dave Leach, Programme Manager
Clara Molina, Programme Coordinator
Peter Foltz, Content Lead, Reading Literacy
Rick Wilmeth, Content Lead, Mathematical Literacy
Michaela Viering, Content Lead, Scientific Literacy
Mark Robeck, Editorial Management

ANNEX B. KEY ELEMENTS OF EACH OF THE SIX STAGES OF DEVELOPMENT

Prenatal period

The Prosperity Outcomes at this stage are a healthy pregnancy and delivery. The Foundations for Success include four family factors: nutrition, no exposure to toxins, the mother's physical health and the mother's emotional health. A number of studies have shown that poor nutrition during the prenatal period modifies the development of the unborn child, leading to low birth weight and a greater susceptibility to coronary heart disease, obesity and diabetes later in life (Barker, 1994; Barker and Sultan, 1995). The exposure of the foetus to environmental toxins or to alcohol or drugs can also compromise the healthy development of the unborn child (Nelson and Panksepp, 1998). The mental and physical well-being of the mother also plays a key role (Liu et al., 1997). The prevalence of healthy pregnancies and deliveries can be increased through the provision of prenatal care and primary health care.

At this stage and in subsequent stages, the framework includes social capital and resources. The term "social capital" is intended to capture positive socializing forces, such as trust amongst community members, social ties and networks connecting people, the presence of strong adult role models, and actively engaged citizens.

Early development (birth to age 2)

The Prosperity Outcomes at this stage include language, cognitive and physical development. The key family foundations include: breast-feeding and nutrition, the mother's physical and emotional health, parenting skills and intra-family relations. These factors can be supported with positive post-natal care and primary health care.

Pre-primary (ages 3 to 5)

The Prosperity Outcomes at this stage includes skills in five domains: awareness of self and environment, cognitive development, language and communication, physical development, and social skills and approaches to learning. These outcomes are consistent with frameworks set out by UNICEF and the United States Congress (Shepard, Kagan and Wurtz, 1998).

Three of the most important family factors affecting children's development in these domains are family involvement, especially reading to the child, positive intra-family relations and parenting styles (Willms, 2002). Children whose parents adopt an "authoritative" parenting style, which incorporates being responsive to the child's needs but also involves setting expectations for positive behaviour, tend to have better outcomes in these domains (Tramonte, Willms and Gauthier, 2013).

Attendance in high-quality early childhood and care (ECEC) programmes has positive short-term outcomes and enduring long-term benefits (Burchinal, Howes and Kontos, 2002; Currie, 2001; Peisner-Feinberg *et al.*, 2001; Ramey and Ramey, 1998; Schweinhart and Weikart, 1998), especially for children from less advantaged backgrounds (Burchinal, Peisner-Feinberg, Bryant and Clifford, 2000). Several large national studies and many smaller local studies suggest that high-quality child care experiences are related, albeit modestly, to child outcomes, even after adjusting for factors such as socioeconomic status (SES) and parental child-rearing attitudes and practices (Howes, Phillips and Whitebook, 1992; Peisner-Feinberg and

Burchinal, 1997; Zill, 1999). Interventions are more effective when early learning programmes take place within a general framework of anti-poverty and community development (Kagan and Zigler, 1987) and when programmes promote family engagement alongside high-quality learning experiences for children (Bertram, Pascel, Bokhari, Gasper and Holtermann, 2002). Programmes for children should be intensive, year-long and conducted by appropriately trained professionals (Leseman, 2002).

Early primary (ages 6 to 9)

After children enter school, there is another critical period that has a dramatic effect on their attainment and performance at age 15. The timely transition from learning-to-read to reading-to-learn, which for most children occurs at about age 8 or 9, is essential to academic success, school attainment and well-being (Snow, Burns and Griffin, 1998). During the primary school years, from kindergarten to Grade 3, considerable emphasis is placed on the development of reading skills. Of course, children learn subject-matter content and acquire a wide range of skills while they are learning to read. But after Grade 3 there is a tacit assumption that children can read fluently and comprehend curricular materials in subject domains such as health, social studies and science. The curriculum changes: students are expected to learn the languages of subject domains and use that language to think critically, solve problems and create new knowledge. The demands for strong reading skills increase as students make their way into the higher grades. Students who lack fundamental reading skills fall further and further behind.

Late primary and lower secondary (ages 10 to 15)

After age 10, during the late primary and lower secondary years, the relationship between early reading skills and future literacy skills is solidified (Francis, Shaywitz, Stuebing, Shaywitz and Fletcher, 1996; Kozminsky and Kozminsky, 2001), as is the relationship between early literacy and social and emotional outcomes (Coleman and Vaughn, 2000). This is the “reading-to-learn” period, during which students require strong literacy skills in all subjects in order to make inferences, monitor comprehension and use higher-order skills, such as previewing, predicting and summarising (O’Reilly and McNamara, 2007). Students who have not made the transition from learning-to-read to reading-to-learn are unable to handle the demands for understanding increasingly complex subject-matter content (Morris, Bloodgood and Perney, 2003).

Upper secondary (ages 16 to 18)

Completing secondary school is a key outcome at this stage. Longitudinal studies that have followed students through to the school-leaving age have identified a number of demographic and school-related factors related to completion (Barrington and Hendricks, 1989; Crane, 1991; Ensminger and Slusarcick, 1992; Fagan and Pabon, 1990; Gilbert, Barr, Clark, Blue and Sunter, 1993; Janosz, LeBlanc, Boulerice and Tremblay, 1997; Rumberger, 1983; Wehlage and Rutter, 1986). Literacy skills, grade repetition, attendance, engagement and positive behaviours are among the most important determinants, and nearly all studies emphasise the role of family socioeconomic status and parental engagement.

ANNEX C. BIOLOGICAL EMBEDDING, CUMULATIVE DEVELOPMENT AND INSTITUTIONAL SELECTION

Biological embedding

Children's potential for success at school is affected by factors during the prenatal period that contribute to a healthy pregnancy and a healthy delivery. Recent advances in neurobiology, molecular biology and genomics have provided compelling evidence that children's early experiences interact with their genetic disposition in ways that affect brain development as well as other neurological and biological systems associated with healthy child development (Boyce, Sokolowski and Robinson, 2012). Some of these biological processes are "biologically embedded" during the prenatal period through epigenetic processes in which chemical signatures are attached to genes that predispose the child to either vulnerability or resilience (Boyce and Kobor, 2015).

At birth, children have billions of neurons and during the course of early development the neurons form connections called synapses in response to environmental stimuli. As this occurs, many of the neurons that are not being used are pruned away. This process of synapse formation and neuron pruning – the sculpting of the brain – is more rapid during certain *critical periods* of the first two or three years of life (McEwan and Schmeck; 1994; Cynader and Frost, 1999). The notion that children's early experiences are biologically embedded is gaining further support from research showing that the development of children's endocrine and immune systems are also influenced by children's environments during the early years (Barr, Beek and Calinoiu, 1999; Gunnar, 1998; McEwan, 1998).

Cumulative development

Children's development and success is cumulative. For example, children develop their literacy skills in a cumulative way as they move from one stage to the next. The rate at which they develop these skills depends on the strength and duration of their exposure to the family, institution and community factors that comprise the Foundations for Success in the Educational Prosperity model. For example, a child's literacy skills at age 15 depends on his or her literacy skills at age 8, which is strongly affected by the quality of instruction the child received during the primary grades. The increase in the child's literacy skills from ages 9 to 18 depends on the quality of instruction he or she received during the late primary and secondary school years.

Institutional selection

When students are successful at one stage of development, their life-course may be altered if they are selected into certain classes, school programmes or schools. For example, children who have strong reading and language skills are more likely to be streamed into classes or school programmes where they benefit from positive peer interactions, a higher quality of instruction and other factors that enable them to develop their skills at a faster pace. On the other hand, children who experience learning difficulties at a particular stage are more likely to be streamed into lower ability classes and have less access to the factors that improve their skills.

ANNEX D. THE EDUCATIONAL PROSPERITY PATHWAY

Figure D.1 shows the effects associated with each of the four processes along the Educational Prosperity pathway. The outcomes at birth are affected by the Foundations for Success (green arrows), which to some extent are biologically embedded (orange arrows) through epigenetic processes. The age 2 outcomes are determined by a cumulative effect (purple arrows) and the Foundations for Success associated with that stage, which include a foundation effect (green arrows) and an effect that is biologically embedded through the sculpting of the brain during critical periods (orange arrows). We assume there are no institutional selection effects at this stage. The age 5 outcomes are also determined by cumulative effects, foundation effects and prior biologically embedded effects. In addition, there can be an institution selection effect (red arrows) if children's outcomes at age 5 are to some extent determined by their access to preschools with varying quality. The outcomes at age 10 and age 18 are affected by the same factors. We assume that the "hard-wired" effects of biological embedding have diminished, although for some outcomes the process of biological embedding continues through to adolescence.

Figure D.1 Four ways that success accumulates

Biological embedding (orange arrows), Foundations for Success (green arrows), cumulative effects (purple arrows), and institutional selection effects (red arrows)