Programme for **I**nternational **S**tudent **A**ssessment

Assessing Scientific, Reading and Mathematical Literacy

A Framework for PISA 2006

OECD
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Foreword

The OECD Programme for International Student Assessment (PISA), created in 1997, represents a commitment by the governments of OECD member countries to monitor the outcomes of education systems in terms of student achievement, within a common international framework. PISA is, above all, a collaborative effort, bringing together scientific expertise from the participating countries and steered jointly by their governments on the basis of shared, policy-driven interests. Participating countries take responsibility for the project at the policy level. 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 comparative assessment. Through participating in these expert groups, countries ensure that the PISA assessment instruments are internationally valid and take into account the cultural and curricular context of OECD member countries, have strong measurement properties, and place an emphasis on authenticity and educational validity.

PISA 2006 represents a continuation of the data strategy adopted in 1997 by OECD countries. The assessed domains continue to be the same as in 2000 and 2003, however *scientific literacy* is now the major domain and the assessment was carried out using a revised framework. The framework for *reading literacy* remains parallel to the ones used in the 2000 and 2003 assessments and the framework for *mathematical literacy* remains parallel to the ones used in the 2003 assessment and they are respectively presented in the publications *Measuring Student Knowledge and Skills – A New Framework for Assessment* (OECD, 1999) and *The PISA 2003 Assessment Framework – Mathematics, Reading, Science and Problem Solving Knowledge and Skills* (OECD, 2003a).

In a similar way, this new publication presents the guiding principle of the PISA 2006 assessment, which is described in terms of the content that students need to acquire, the processes that need to be performed and the contexts in which knowledge and skills are applied. Further, it illustrates the assessment domains with a range of sample tasks. These have been developed by expert panels under the direction of Raymond Adams, Ross Turner, Barry McCrae and Juliette Mendelovits from the Australian Council for Educational Research (ACER). The science expert group was chaired by Rodger Bybee of the Biological Science Curriculum Study from the United States. The mathematics expert group panel was chaired by Jan de Lange of the University of Utrecht from the Netherlands and the reading expert group was chaired by Irwin Kirsch of Educational Testing Service in the United States until October 2005. After this time John de Jong of the Language Testing Services from the Netherlands became acting chair. The members of the expert groups are listed at the end of this publication. The frameworks have also been reviewed by expert panels in each of the participating countries.

This publication was prepared by the OECD Secretariat, principally by John Cresswell and Sophie Vayssettes. The report is published on the responsibility of the Secretary-General of the OECD.

Table of Contents

Introduction

OVERVIEW

The OECD Programme for International Student Assessment (PISA) is a collaborative effort undertaken by all member countries and a number of non-member partner countries to measure how well students, at age 15, are prepared to meet the challenges they may encounter in future life. Age 15 is chosen because at this age, in most OECD countries, students are approaching the end of compulsory schooling, and so, some measure of the knowledge, skills and attitudes accumulated over approximately ten years of education is gained from an assessment at this time. The PISA assessment takes a broad approach to assessing knowledge, skills and attitudes that reflect current changes in curricula, moving beyond the school based approach towards the use of knowledge in everyday tasks and challenges. The skills acquired reflect the ability of students to continue learning throughout their lives by applying what they learn in school to non-school environments, evaluating their choices and making decisions. The assessment, jointly guided by the participating governments, brings together the policy interests of countries by applying scientific expertise at both national and international levels.

PISA combines the assessment of domain-specific cognitive areas such as science, mathematics and reading with information on students' home background, their approaches to learning, their perceptions of their learning environments and their familiarity with computers. A high priority in PISA 2006 is an innovative assessment of student attitudes towards science – questions about this were contextualised within the cognitive part of the test. Bringing the attitude items closer to the cognitive questions allowed questions to be targeted at specific areas, with the focus on interest in science and students' support for scientific enquiry. Student outcomes are then associated with these background factors.

PISA uses: *i)* strong quality assurance mechanisms for translation, sampling and test administration; *ii)* measures to achieve cultural and linguistic breadth in the assessment materials, particularly through countries' participation in the development and revision processes for the production of the items; and *iii)* state of the art technology and methodology for data handling. The combination of these measures produces high quality instruments and outcomes with superior levels of validity and reliability to improve the understanding of education systems as well as students' knowledge, skills and attitudes.

PISA is based on a dynamic model of lifelong learning in which new knowledge and skills necessary for successful adaptation to a changing world are continuously acquired throughout life. PISA focuses on things that 15-year-old students will need in the future and seeks to assess what they can do with what they have learned. The assessment is informed, but not constrained, by the common denominator of national curricula. Thus, while it does assess students' knowledge, PISA also examines their ability to reflect, and to apply their knowledge and experience to real world issues. For example, in order to understand and evaluate scientific advice on food safety an adult would need not only to know some basic facts about the composition of nutrients, but also to be able to apply that information. The term "literacy" is used to encapsulate this broader concept of knowledge and skills.

Box A • What is PISA?

Basics

- An internationally standardised assessment that was jointly developed by participating countries and administered to 15-year-olds in educational programmes.
- A survey implemented in 43 countries in the first cycle (32 in 2000 and 11 in 2002), 41 countries in the second cycle (2003) and 56 in the third cycle (2006).
- The test is typically administered to between 4 500 and 10 000 students in each country.

Content

- PISA 2006 covers the domains of *reading, mathematical* and *scientific literacy* not so much in terms of mastery of the school curriculum, but in terms of important knowledge and skills needed in adult life.
- Emphasis is on the mastery of processes, the understanding of concepts and the ability to function in various situations within each domain.

Methods

- Paper-and-pencil tests are used, with assessments lasting a total of two hours for each student.
- Test items are a mixture of multiple-choice items and questions requiring students to construct their own responses. The items are organised in groups based on a passage setting out a real-life situation.
- A total of about 390 minutes of test items is covered, with different students taking different combinations of test items.
- Students answer a background questionnaire, which takes 30 minutes to complete, providing information about themselves and their homes. School principals are given a 20-minute questionnaire about their schools.

Assessment cycle

- The assessment takes place every three years with a strategic plan in place extending through to 2015.
- Each of these cycles looks in depth at a major domain, to which two-thirds of testing time is devoted; the other domains provide a summary profile of skills. Major domains are *reading literacy* in 2000, *mathematical literacy* in 2003 and *scientific literacy* in 2006.

Outcomes

- A basic profile of knowledge and skills among 15-year-old students
- Contextual indicators relating results to student and school characteristics, with emphasis in 2006 placed on assessing students' attitudes towards science
- Trend indicators showing how results change over time
- A valuable knowledge base for policy analysis and research

PISA is designed to collect information through three-yearly cycles and presents data on the *reading, mathematical* and *scientific literacy* of students, schools and countries. It provides insights into the factors that influence the development of skills and attitudes at home and at school, and examines how these factors interact and what the implications are for policy development.

This publication presents the conceptual framework underlying the PISA 2006 assessments, including a re-developed and expanded framework for *scientific literacy*, incorporating an innovative component on the assessment of students' attitudes towards science, and the frameworks for the assessment of reading and mathematics. Within each domain, the framework defines the contents that students need to acquire, the processes that need to be performed and the contexts in which knowledge and skills are applied. Finally, it illustrates the domain and their aspects with sample tasks.

BASIC FEATURES OF PISA 2006

PISA 2006 is the third cycle of a data strategy defined in 1997 by participating countries. The publications *Measuring Student Knowledge and Skills – A New Framework for Assessment* (OECD, 1999) and *The PISA 2003 Assessment Framework – Mathematics, Reading, Science and Problem Solving Knowledge and Skills* (OECD, 2003a) presented the conceptual framework underlying the first two cycles of PISA. The results from those cycles were presented in the publications *Knowledge and Skills for Life – First Results from PISA 2000* (OECD, 2001) and *Learning for Tomorrow's World: First Results from PISA 2003* (OECD, 2004), and are also available on the PISA website: *www.pisa.oecd.org*. The results allow national policy makers to compare the performance of their education systems with those of other countries. Similar to the previous cycles, the 2006 assessment covers the domains of *reading, mathematical* and *scientific literacy*, with the major focus on *scientific literacy*. Students also respond to a background questionnaire, and additional supporting information is gathered from the school authorities. Fifty-six countries and regions, including all 30 OECD member countries, are taking part in the PISA 2006 assessment. Together, they comprise almost 90% of the world's economy.

Since the aim of PISA is to assess the cumulative yield of education systems at an age where compulsory schooling is still largely universal, testing focused on 15-year-olds enrolled in both school-based and work-based educational programmes. Between 5 000 and 10 000 students from at least 150 schools will typically be tested in each country, providing a good sampling base from which to break down the results according to a range of student characteristics.

The primary aim of the PISA assessment is to determine the extent to which young people have acquired the wider knowledge and skills in *reading, mathematical* and *scientific literacy* that they will need in adult life. The assessment of cross-curricular competencies continues to be an integral part of PISA 2006. The main reasons for this broadly oriented approach are:

- Although specific knowledge acquisition is important in school learning, the application of that knowledge in adult life depends crucially on the acquisition of broader concepts and skills. In science, having specific knowledge, such as the names of plants and animals, is of less value than understanding broad topics such as energy consumption, biodiversity and human health in thinking about the issues under debate in the adult community. In reading, the capacity to develop interpretations of written material and to reflect on the content and qualities of text are central skills. In mathematics, being able to reason quantitatively and to represent relationships or

dependencies is more apt than the ability to answer familiar textbook questions when it comes to deploying mathematical skills in everyday life.

- In an international setting, a focus on curriculum content would restrict attention to curriculum elements common to all or most countries. This would force many compromises and result in an assessment too narrow to be of value for governments wishing to learn about the strengths and innovations in the education systems of other countries.
- Certain broad, general skills are essential for students to develop. They include communication, adaptability, flexibility, problem solving and the use of information technologies. These skills are developed across the curriculum and an assessment of them requires a broad cross-curricular focus.

PISA is not a single cross-national assessment of the reading, mathematics and science skills of 15-year-old students. It 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. On each occasion, one domain will be tested in detail, taking up nearly two-thirds of the total testing time. The major domain was *reading literacy* in 2000 and *mathematical literacy* in 2003, and is *scientific literacy* in 2006. This will provide a thorough analysis of achievement in each area every nine years and a trend analysis every three.

Similar to previous cycles of PISA, the total time spent on the PISA 2006 tests by each student is two hours, but information is obtained on about 390 minutes worth of test items. The total set of questions is packaged into 13 linked testing booklets. Each booklet is taken 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 sub-groups within a country (such as males and females, and students from different social and economic contexts). Students also spend 30 minutes answering questions for the context questionnaire.

The PISA assessment provides three main types of outcomes:

- Basic indicators that provide baseline profile of the knowledge and skills of students.
- Contextual indicators that show how such skills relate to important demographic, social, economic and educational variables.
- Indicators on trends that emerge from the on-going nature of the data collection and that show changes in outcome levels and distributions, and in relationships between student-level and school-level background variables and outcomes.

Although indicators are an adequate means of drawing attention to important issues, they are not usually capable of providing answers to policy questions. PISA has therefore also developed a policy-oriented analysis plan that will go beyond the reporting of indicators.

WHAT MAKES PISA UNIQUE

PISA is not the first international comparative survey of student achievement. Others have been conducted over the past 40 years, primarily developed by the International Association for the Evaluation of Educational Achievement (IEA) and by the Education Testing Service's International Assessment of Educational Progress (IAEP).

More importantly, these surveys have concentrated on outcomes linked directly to the curriculum and then only to those parts of the curriculum that are essentially common across the participating countries. Aspects of the curriculum unique to one country or a small number of countries have usually not been taken into account in the assessments.

PISA takes a different approach in a number of respects:

- Its *origin:* an initiative taken by governments, whose policy interests the results are addressing.
- Its *regularity:* the commitment to cover multiple assessment domains with updates every three years makes it possible for countries to monitor regularly and predictably their progress in meeting key learning objectives.
- The *age-group covered:* assessing young people near the end of their compulsory schooling gives a useful indication of the performance of education systems. While most young people in OECD countries continue their initial education beyond the age of 15, this is normally close to the end of the initial period of basic schooling in which all young people follow a broadly common curriculum. It is useful to determine, at that stage, the extent to which they have acquired knowledge and skills that will help them in the future, including the individualised paths of further learning they may follow.
- The *knowledge and skills tested:* these are defined not primarily in terms of a common denominator of national school curricula but in terms of what skills are deemed to be essential for future life. This is the most fundamental feature of PISA. School curricula are traditionally constructed largely in terms of bodies of information and techniques to be mastered. They traditionally focus less, within curriculum areas, on the skills to be developed in each domain for use generally in adult life. They focus even less on more general competencies, developed across the curriculum, to solve problems and apply ideas and understanding to situations encountered in life. PISA does not exclude curriculum-based knowledge and understanding, but it tests for it mainly in terms of the acquisition of broad concepts and skills that allow knowledge to be applied. Further, PISA is not constrained by the common denominator of what has been specifically taught in the schools of participating countries.

This emphasis on testing in terms of mastery and broad concepts is particularly significant in light of the concern among nations to develop human capital, which the OECD defines as:

> The knowledge, skills, competencies and other attributes embodied in individuals that are relevant to personal, social and economic well-being.

Estimates of human capital have tended, at best, to be derived using proxies such as level of education completed. When the interest in human capital is extended to include attributes that permit full social and democratic participation in adult life and that equip people to become lifelong learners, the inadequacy of these proxies becomes even clearer.

By directly testing for knowledge and skills close to the end of basic schooling, PISA examines the degree of preparedness of young people for adult life and, to some extent, the effectiveness of education systems. Its ambition is to assess achievement in relation to the underlying objectives (as defined by society) of education systems, not in relation to the teaching and learning of a body of knowledge. This view of educational outcomes is needed if schools and education systems are to be encouraged to focus on modern challenges.

AN OVERVIEW OF WHAT IS BEING ASSESSED IN EACH DOMAIN

Box B presents a definition of the three domains assessed in PISA 2006. The definitions all emphasise functional knowledge and skills that allow one to participate actively 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 equipped to take part in decision-making processes. In the more complex tasks in PISA, students were asked to reflect on and evaluate material, not just to answer questions that have single correct answers.

Box B • Definitions of the domains

Scientific literacy: An individual's scientific knowledge and use of that knowledge to identify questions, to acquire new knowledge, to explain scientific phenomena, and to 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, and willingness to engage in science-related issues, and with the ideas of science, as a reflective citizen.

Reading literacy: An individual's capacity to understand, use and reflect on 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 identify and understand the role that mathematics plays in the world, to make well-founded judgements and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen.

Scientific literacy (elaborated in Chapter 1) is defined as the ability to use scientific knowledge and processes not only to understand the natural world but to participate in decisions that affect it. *Scientific literacy* is assessed in relation to:

- *Scientific knowledge or concepts:* These constitute the links that aid understanding of related phenomena. In PISA, while the concepts are the familiar ones relating to physics, chemistry, biological sciences and earth and space sciences, they are applied to the content of the items and not just recalled.
- *Scientific processes:* These are centred on the ability to acquire, interpret and act upon evidence. Three such processes present in PISA relate to: *i)* describing, explaining and predicting scientific phenomena, *ii)* understanding scientific investigation, and *iii)* interpreting scientific evidence and conclusions.
- *Situations or contexts:* These concern the application of scientific knowledge and the use of scientific processes. The framework identifies three main areas: science in life and health, science in Earth and environment, and science in technology.

Reading literacy (elaborated in Chapter 2) is defined in terms of students' ability to understand, use and reflect on written text to achieve their purposes. This aspect of literacy has been well established by previous surveys such as the International Adult Literacy Survey (IALS), but is taken further in PISA by the introduction of an active element – the capacity not just to understand

a text but to reflect on it, drawing on one's own thoughts and experiences. *Reading literacy* is assessed in relation to the:

- *Text format:* Often students' reading assessments have focused on *continuous texts* or prose organised in sentences and paragraphs. PISA introduces in addition *non-continuous texts* that present information in other ways, such as in lists, forms, graphs, or diagrams. It will also distinguish between a range of prose forms, such as narration, exposition and argumentation. These distinctions are based on the principle that individuals will encounter a range of written material in their work-related adult life (*e.g.* application, forms, advertisements) and that it is not sufficient to be able to read a limited number of types of text typically encountered in school.
- *Reading processes (aspects):* Students are not assessed on the most basic reading skills, as it is assumed that most 15-year-old students will have acquired these. Rather, they are expected to demonstrate their proficiency in retrieving information, forming a broad general understanding of the text, interpreting it, reflecting on its contents and reflecting on its form and features.
- *Situations:* These are defined by the use for which the text was constructed. For example, a novel, personal letter or biography is written for people's personal use; official documents or announcements for public use; a manual or report for occupational use; and a textbook or worksheet for educational use. Since some groups may perform better in one reading situation than in another, it is desirable to include a range of types of reading in the assessment items.

Mathematical literacy (elaborated in Chapter 3) is concerned with the ability of students to analyse, reason, and communicate ideas effectively as they pose, formulate, solve, and interpret solutions to mathematical problems in a variety of situations. *Mathematical literacy* is assessed in relation to the:

- *Mathematical content:* This is defined mainly in terms of four overarching ideas (*quantity*, *space and shape*, *change and relationships,* and *uncertainty*) and only secondarily in relation to curricular strands such as numbers, algebra and geometry.
- *Mathematical processes:* These are defined by general mathematical competencies. These include the use of mathematical language, modelling and problem-solving skills. Such skills, however, are not separated out in different test items, since it is assumed that a range of competencies will be needed to perform any given mathematical task. Rather, questions are organised in terms of competency clusters defining the type of thinking skill needed.
- *Situations:* These are defined in terms of the ones in which mathematics is used, based on their distance from the students. The framework identifies five situations: personal, educational, occupational, public and scientific.

ASSESSING AND REPORTING PISA 2006

Similar to the previous assessments in PISA, the assessment in 2006 consists of pencil and paper instruments for reasons of feasibility. The assessment includes a variety of types of questions. Some require students to select or produce simple responses that can be directly compared with a single correct answer, such as multiple-choice or closed-constructed response items. These questions have either a correct or incorrect answer and often assess lower-order skills. Others are more constructive, requiring students to develop their own responses designed to measure broader constructs than those captured by more traditional surveys, allowing for a wider range of acceptable responses and more complex marking that can include partially correct responses.

Not all students answer all questions in the assessment. The PISA 2006 test units are arranged in 13 clusters, with each cluster designed to occupy 30 minutes of testing time. There are seven science clusters, two reading clusters and four mathematics clusters. The clusters are placed in 13 booklets, according to a rotated test design. Each booklet contains four clusters and each student is assigned one of these two-hour booklets. There is at least one science cluster in each booklet.

Literacy in PISA is assessed through units consisting of a stimulus (*e.g.* text, table, chart, figures etc.) followed by a number of tasks associated with this common stimulus. This is an important feature, allowing questions to go into greater depth than they could if each question introduced a wholly new context. It allows time for the student to digest material that can then be used to assess multiple aspects of performance.

Results from PISA have been reported using scales with an average score of 500 and a standard deviation of 100 for all three domains, which means that two-thirds of students across OECD countries scored between 400 and 600 points. These scores represent degrees of proficiency in a particular aspect of literacy. As *reading literacy* was the major domain in 2000, the reading scales were divided into five levels of knowledge and skills. The main advantage of this approach is that it describes what students can do by associating the tasks with levels of difficulty. Additionally, results were also presented through three subscales of reading: retrieving information, interpreting texts, and reflection and evaluation. A proficiency scale was also available for *mathematical* and *scientific literacy*, though without levels thus recognising the limitation of the data from minor domains. PISA 2003 built upon this approach by specifying six proficiency levels for the mathematical literacy scale, following a similar approach to what was done in reading. There were four subscales in *mathematical literacy: space and shape, change and relationships, quantity* and *uncertainty*. The reporting of *scientific literacy* will be in a similar manner and will also present results in different areas. PISA 2003 offered the first opportunity to present trend results for *reading, mathematical* and *scientific literacy* and the results from PISA 2006 will provide extra information for this analysis.

THE CONTEXT QUESTIONNAIRES AND THEIR USE

To gather contextual information, PISA asks students and the principals of their schools to respond to background questionnaires of around 30 minutes in length. These questionnaires are central to the analysis of results in terms of a range of student and school characteristics. The questionnaires from PISA 2000 and 2003 are available on the PISA website: *www.pisa.oecd.org.*

The questionnaires seek information about:

- Students and their family backgrounds, including their economic, social and cultural capital
- Aspects of students' lives, such as their attitudes towards learning, their habits and life inside school, and their family environment
- Aspects of schools, such as the quality of the schools' human and material resources, public and private control and funding, decision-making processes, and staffing practices
- Context of instruction, including institutional structures and types, class size, and the level of parental involvement

- Strategies of self-regulated learning, motivational preferences and goal orientations, self-related cognition mechanisms, action control strategies, preferences for different types of learning situations, learning styles, and social skills required for co-operative or competitive learning
- Aspects of learning and instruction in science, including students' motivation, engagement and confidence with science, and the impact of learning strategies on achievement related to the teaching and learning of science

Two additional questionnaires are offered as international options:

- A computer familiarity questionnaire focusing on: *i)* availability and use of information and communications technology (ICT), including the location where ICT is mostly used as well as the type of use; *ii)* ICT confidence and attitudes, including self-efficacy and attitudes towards computers; and *iii)* learning background of ICT, focusing on where students learned to use computers and the Internet. The OECD published a report resulting from analysis of data collected via this questionnaire in 2003, *Are Students Ready for a Technology-Rich World? What PISA Studies Tell Us* (OECD, 2005).
- A parent questionnaire focusing on a number of topics including the student's past science activities, parents' views on the student's school, parents' views on science in the student's intended career and the need for scientific knowledge and skills in the job market, parents' views on science and the environment, the cost of education services, and parents' education and occupation.

The contextual information collected through the student and school questionnaires, as well as the optional computer familiarity and parent questionnaires, comprises only a part of the total amount of information available to PISA. Indicators describing the general structure of the education systems (their demographic and economic contexts – for example, costs, enrolments, school and teacher characteristics, and some classroom processes) and their effect on labour market outcomes are already routinely developed and applied by the OECD.

COLLABORATIVE DEVELOPMENT OF PISA AND ITS ASSESSMENT FRAMEWORKS

PISA represents a collaborative effort among the OECD member governments to provide a new kind of assessment of student achievement on a recurring basis. The assessments are developed co-operatively, agreed by participating countries, and implemented by national organisations. The constructive co-operation of students, teachers and principals in participating schools has been crucial to the success of PISA during all stages of the development and implementation.

The PISA Governing Board (PGB), representing all nations at the 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. This includes setting priorities for the development of indicators, for the establishment of the assessment instruments and for the reporting of the results. Experts from participating countries also serve on working groups charged with linking the PISA policy objectives with the best internationally available technical expertise in the different assessment domains. By participating in these expert groups, countries ensure that the instruments are internationally valid and take into account the cultural and educational contexts in OECD member countries. They also ensure that the assessment materials have strong measurement properties and that the instruments emphasise authenticity and educational validity.

Participating countries implement PISA at the national level, through National Project Managers (NPM), subject to the agreed administration procedures. National Project Managers play a vital role in ensuring that implementation is of high quality. They also verify and evaluate the survey results, analyses, reports and publications.

The design and implementation of the present surveys, within the framework established by the PGB, is the responsibility of an international consortium led by the Australian Council for Educational Research (ACER). Other partners in this consortium include the National Institute for Educational Measurement (CITO) in the Netherlands, WESTAT and the Educational Testing Service (ETS) in the United States, and the National Institute for Educational Policy Research (NIER) in Japan.

The OECD Secretariat has overall managerial responsibility for the programme, monitors its implementation on a day-to-day basis, acts as the secretariat for the PGB, builds consensus among countries and serves as the interlocutor between the PGB and the international consortium charged with implementation. The OECD Secretariat is also responsible for the production of the indicators, and the analysis and preparation of the international reports and publications in co-operation with the PISA consortium, in close consultation with member countries both at the policy level (PGB) and at the implementation level (National Project Managers).

The development of the PISA frameworks has been a continuous effort since the programme was created in 1997 and can be described as a sequence:

- Development of a working definition for the assessment domain and description of the assumptions that underlie that definition
- Evaluation of how to organise the tasks constructed in order to report to policy makers and researchers on student achievement in the domain, and identification of key characteristics that should be taken into account when constructing assessment tasks for international use
- Operationalisation of key characteristics used in test construction, with definitions based on existing literature and experience in conducting other large-scale assessments
- Validation of the variables and assessment of the contribution each makes to understanding task difficulty across the various participating countries
- Preparation of an interpretative scheme for the results

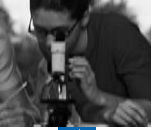
While the main benefit of constructing and validating a framework for each of the domains is improved measurement, there are other potential benefits:

- A framework provides a common language and a vehicle for discussing the purpose of the assessment and what it is trying to measure. Such a discussion encourages the development of a consensus around the framework and the measurement goals.
- An analysis of the kinds of knowledge and skills associated with successful performance provides a basis for establishing standards or levels of proficiency. As the understanding of what is being measured and the ability to interpret scores along a particular scale evolve, an empirical basis for communicating a richer body of information to various constituencies can be developed.
- Identifying and understanding particular variables that underlie successful performance further the ability to evaluate what is being measured and to make changes to the assessment over time.

The understanding of what is being measured and its connection to what we say about students provides an important link between public policy, assessment and research which, in turn, enhances the usefulness of the data collected.

1

Scientific Literacy



INTRODUCTION

The assessment of *scientific literacy* has particular importance in PISA 2006, where it is the major domain being assessed. Since this is the first time that it is being tested in such detail, the domain has undergone considerable development since the 2003 survey, with an expanded interpretation of what is being assessed. This involves not only a more detailed description of *scientific literacy,* but also an important innovation in the approach to assessment that has relevance for all of PISA in the future. For the first time, the main assessment instrument includes questions on attitudes alongside the testing of cognitive abilities and knowledge. By exploring the extent to which the issues that they are addressing in the course of the test excite students' interest, this strengthens the assessment of the attitudinal and motivational characteristics that will be important to their future engagement with science. Previously, questions about these aspects have been limited to a separate questionnaire asking in more general terms about aspects such as interest and motivation.

An understanding of science and technology is central to a young person's preparedness for life in modern society. It enables an individual to participate fully in a society in which science and technology play a significant role. This understanding also empowers individuals to participate appropriately in the determination of public policy where issues of science and technology impact on their lives. An understanding of science and technology contributes significantly to the personal, social, professional and cultural lives of all people.

A large proportion of the situations, problems and issues encountered by individuals in their daily lives require some understanding of science and technology before they can be fully understood or addressed. Science and technology related issues confront individuals at personal, community, national and even global levels. Therefore, national leaders should be encouraged to ask about the degree to which all individuals in their respective countries are prepared to deal with these issues. A critical aspect of this is how young people respond to scientific questions when they emerge from school. An assessment at age 15 provides an early indication of how they may respond in later life to the diverse array of situations that involve science and technology.

As the basis for an international assessment of 15-year-old students, it seems reasonable, therefore, to ask: "What is it important for citizens to know, value, and be able to do in situations involving science and technology?" Answering this question establishes the basis for an assessment of students in these respects: their knowledge, values and abilities today relate to what is needed in the future. Central to the answer are the competencies that lie at the heart of the PISA 2006 definition of *scientific literacy*. These ask how well students:

- Identify scientific issues
- Explain phenomena scientifically
- Use scientific evidence

These competencies require students to demonstrate, on the one hand, knowledge, cognitive abilities, and on the other, attitudes, values and motivations as they meet and respond to science-related issues.

The issue of identifying what citizens should know, value and be able to do in situations involving science and technology, seems simple and direct. However doing so raises questions about scientific

understanding, and does not imply mastery of all scientific knowledge. This framework is guided by reference to what citizens require. As citizens, what knowledge is most appropriate? An answer to this question certainly includes basic concepts of the science disciplines, but that knowledge must be used in contexts that individuals encounter in life. In addition, people often encounter situations that require some understanding of science as a process that produces knowledge and proposes explanations about the natural world.[1] Further, they should be aware of the complementary relationships between science and technology, and how science-based technologies pervade and influence the nature of modern life.

What is important for citizens to value about science and technology? An answer should include the role and contributions to society of science and of science-based technology, and their importance in many personal, social, and global contexts. Accordingly, it seems reasonable to expect individuals to have an interest in science, to support the process of scientific enquiry and to act responsibly towards natural resources and the environment.

What is important for individuals to be able to do that is science related? People often have to draw appropriate conclusions from evidence and information given to them; they have to evaluate claims made by others on the basis of the evidence put forward and they have to distinguish personal opinion from evidence-based statements. Often the evidence involved is scientific, but science has a more general role to play as well since it is concerned with rationality in testing ideas and theories against evidence. Of course this does not deny that science includes creativity and imagination, attributes that have always played a central part in advancing human understanding of the world.

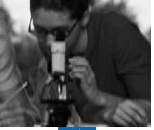
Can citizens distinguish claims that are scientifically sound from those that are not? Ordinary citizens are generally not called on to judge the worth of major theories or potential advances in science. But they do make decisions based on the facts in advertisements, evidence in legal matters, information about their health and issues concerning local environments and natural resources. An educated person should be able to distinguish the kinds of questions that can be answered by scientists and the kinds of problems that can be solved by science-based technologies from those that cannot be answered in these ways.

DEFINITION OF THE DOMAIN

Current thinking about the desired outcomes of science education emphasises scientific knowledge (including knowledge of the scientific approach to enquiry) and an appreciation of science's contribution to society. These outcomes require an understanding of important concepts and explanations of science, and of the strengths and limitations of science in the world. They imply a critical stance and a reflective approach to science (Millar and Osborne, 1998).

Such goals provide an orientation and emphasis for the science education of all people (Fensham, 1985). The competencies assessed in PISA 2006 are broad and include aspects that relate to personal utility, social responsibility, and the intrinsic and extrinsic value of scientific knowledge.

The above discussion frames a central point of the PISA 2006 science assessment: The assessment should focus on competencies that clarify what 15-year-old students know, value and are able to do within reasonable and appropriate personal, social and global contexts. This perspective differs from one grounded exclusively in school science programmes and extensively based only on the disciplines



of science; but it includes problems situated in educational contexts and also in professional ones, and recognises the essential place of the knowledge, methods, attitudes, and values that define scientific disciplines. The term that best describes the overall purposes of the PISA 2006 science assessment is *scientific literacy* (Bybee, 1997b; Fensham, 2000; Graber and Bolte, 1997; Mayer, 2002; Roberts, 1983; UNESCO, 1993).

PISA 2006 aims to assess both the cognitive and affective aspects of students' *scientific literacy*. The cognitive aspects include students' knowledge and their capacity to use this knowledge effectively, as they carry out certain cognitive processes that are characteristic of science and scientific enquiries of personal, social, or global relevance. In assessing scientific competencies, PISA is concerned with issues to which scientific knowledge can contribute and which will involve students, either now or in the future, in making decisions. From the point of view of their scientific competencies, students respond to such issues in terms of their understanding of relevant scientific knowledge, their ability to access and evaluate information, their ability to interpret evidence bearing on the issue and their ability to identify the scientific and technological aspects of the issue (Koballa, Kemp and Evans, 1997; Law, 2002). PISA also assesses non-cognitive aspects: how students respond affectively. Attitudinal aspects of their response engage their interest, sustain their support, and motivate them to take action (Schibeci, 1984). Through such considerations we are led to define the overarching domain of *scientific literacy* for PISA 2006.

Box 1.1 • **Scientific knowledge: PISA 2006 terminology**

The term "scientific knowledge" is used throughout this framework to refer collectively to both *knowledge of science* and *knowledge about science*. *Knowledge of science* refers to knowledge of the natural world across the major fields of physics, chemistry, biological science, Earth and space science, and science-based technology. *Knowledge about science* refers to knowledge of the means (scientific enquiry) and goals (scientific explanations) of science.

The term *scientific literacy* has been chosen because it is recognised as representing the goals of science education that should apply to all students, connotes a broadness and an applied nature to the purposes of science education, represents a continuum of scientific knowledge and the cognitive abilities associated with scientific enquiry, incorporates multiple dimensions, and includes the relationships between science and technology. Together, the scientific competencies at the heart of the definition characterise a foundation for *scientific literacy,* and the objective of the PISA 2006 science assessment – to assess the degree to which the competencies have been developed (Bybee, 1997a; Fensham, 2000; Law, 2002; Mayer and Kumano, 2002).

> **Box 1.2 • PISA 2006 *Scientific literacy***
>
> For the purposes of PISA 2006, *scientific literacy*[2] 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

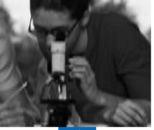
The following remarks further clarify this definition.

Scientific literacy

Using the term "scientific literacy" rather than "science" underscores the importance that the PISA 2006 science assessment places on the application of scientific knowledge in the context of life situations, compared with the simple reproduction of traditional school science knowledge. The functional use of knowledge requires the application of those processes that are characteristic of science and scientific enquiry (the scientific competencies) and is regulated by the individual's appreciation, interest, values, and action relative to scientific matters. A student's ability to carry out the scientific competencies involves both knowledge of science and an understanding of the characteristics of science as a way of acquiring knowledge (*i.e.* knowledge about science). The definition also recognises that the disposition to carry out these competencies depends upon an individual's attitudes towards science and a willingness to engage in science-related issues. Note that non-cognitive aspects such as motivation are themselves considered to be competencies.

Knowledge and use of that knowledge to identify questions, to acquire new knowledge, to explain scientific phenomena and to draw evidence-based conclusions

Knowledge for this definition of *scientific literacy* implies far more than the ability to recall information, facts, and names. The definition includes knowledge of science (knowledge about the natural world) and knowledge about science itself. The former includes understanding fundamental scientific concepts and theories; the latter includes understanding the nature of science as a human activity and the power and limitations of scientific knowledge. The questions to be identified are those that can be answered by scientific enquiry, again requiring knowledge about science as well as scientific knowledge of the specific topics involved. Of significant note for the definition of *scientific literacy* is the fact that individuals must often acquire knowledge that is new to them, not through their own scientific investigations, but through resources such as libraries and the internet. Drawing evidence-based conclusions means knowing, selecting and evaluating information and data, while recognising that there is often not sufficient information



to draw definite conclusions, thus making it necessary to speculate, cautiously and consciously, about the information that is available.

Characteristic features of science as a form of human knowledge and enquiry

As expressed here, *scientific literacy* implies that students should have some understanding of how scientists obtain data and propose explanations, recognise key features of scientific investigations, and the types of answers one can reasonably expect from science. For example, scientists use observations and experiments to gather data about objects, organisms, and events in the natural world. The data are used to propose explanations that become public knowledge and may be used in various forms of human activity. Some key features of science include: the collection and use of data – data collection is guided by ideas and concepts (sometimes stated as hypotheses) and includes issues of relevance, context and accuracy; the tentative nature of knowledge claims; an openness to sceptical review; the use of logical arguments; and the obligation to make connections to current and historical knowledge, and to report the methods and procedures used in obtaining evidence.

How science and technology shape our material, intellectual, and cultural environments

The key points in this statement include the idea that science is a human endeavour, one that influences our societies and us as individuals. Further, technological development also is a human endeavour (Fleming, 1989). Although science and technology differ in aspects of their purposes, processes, and products, it is the case that they also are closely related and, in many respects, complementary. In this regard, the definition of *scientific literacy* proposed here includes the nature of science and of technology and their complementary relationships. As individuals we make decisions through public policies that influence the directions of science and technology. Science and technology play paradoxical roles in society as they propose answers to questions and provide solutions to problems, but may also create new questions and problems.

Willingness to engage in science-related issues and with the ideas of science as a reflective citizen

The meanings conveyed in the first part of this statement are wider than taking note and taking action as required; it implies having continuing interest in, having opinions about and participating in current and future science-based issues. The second part of the statement covers various aspects of attitudes and values that individuals may have towards science. The phrase implies a person who has an interest in scientific topics, thinks about science-related issues, has a concern for issues of technology, resources and the environment, and reflects on the importance of science in personal and social perspectives.

Inevitably, *scientific literacy* draws upon *reading* and *mathematical literacies* (Norris and Phillips, 2003). For example, *reading literacy* is necessary when a student is demonstrating an understanding of scientific terminology. Similarly, aspects of *mathematical literacy* are required in data interpretation contexts. The intersection of these other literacies with the PISA 2006 definition and assessment of scientific literacy cannot be avoided; however, at the core of each assessment task there should be aspects that are unambiguously *scientific literacy*.

Compared to the definition of *scientific literacy* for PISA in 2000 and 2003, the definition for 2006 has been elaborated and enhanced. For the previous two assessments, when science was a minor domain, *scientific literacy* was defined as follows:

> *Scientific literacy* is 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, 1999, 2000, 2003a)

The initial assertions of the 2000, 2003 and 2006 definitions are fundamentally the same in that they centre on individuals' uses of scientific knowledge to draw conclusions. While the 2000 and 2003 definition embedded knowledge of science and understandings about science within the terms of scientific knowledge, the 2006 definition separates and elaborates this aspect of *scientific literacy* through the addition of terms that underscore students' knowledge about the characteristic features of science. Both definitions then refer to the application of scientific knowledge to understand, and ultimately to make informed decisions about, the natural world. In PISA 2006, this part of the definition is enhanced by the addition of knowledge of the relationship between science and technology – an aspect of *scientific literacy* that was assumed but not elaborated in the earlier definition. In today's world, science and technology are closely linked, often having synergistic relationships with each other.

In contrast to the earlier definition, the PISA 2006 definition of *scientific literacy* has been expanded by explicitly including attitudinal aspects of students' responses to issues of scientific and technological relevance. In summary, the 2006 definition is conceptually in accord with the 2000 and 2003 definition, with the exception of the addition of attitudinal responses. However the attitudinal element is reported separately and therefore does not impact on the comparability of the cognitive aspect over time. Other changes, for example elaborating knowledge about science, and science-based technology, represent an increased emphasis on particular aspects that were embedded or assumed in the earlier definition.

ORGANISATION OF THE DOMAIN

The definition of *scientific literacy* proposed here provides for a continuum from less developed to more developed *scientific literacy* – that is, individuals are deemed to be more or less scientifically literate; they are not regarded as either scientifically literate or scientifically illiterate (Bybee, 1997a and 1997b). So, for example, the student with less developed *scientific literacy* might be able to recall simple scientific factual knowledge and to use common scientific knowledge in drawing or evaluating conclusions. A student with more developed *scientific literacy* will demonstrate the ability to create and use conceptual models to make predictions and give explanations, analyse scientific investigations, relate data as evidence, evaluate alternative explanations of the same phenomena, and communicate conclusions with precision.

For purposes of assessment, the PISA 2006 definition of *scientific literacy* may be characterised as consisting of four interrelated aspects:

- *Context:* recognising life situations involving science and technology.
- *Knowledge:* understanding the natural world on the basis of scientific knowledge that includes both knowledge of the natural world, and knowledge about science itself.
- *Competencies:* demonstrating competencies that include identifying scientific issues, explaining phenomena scientifically, and drawing conclusions based on evidence.
- *Attitudes:* indicating an interest in science, support for scientific enquiry, and motivation to act responsibly towards, for example, natural resources and environments.

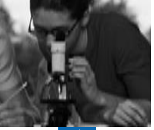
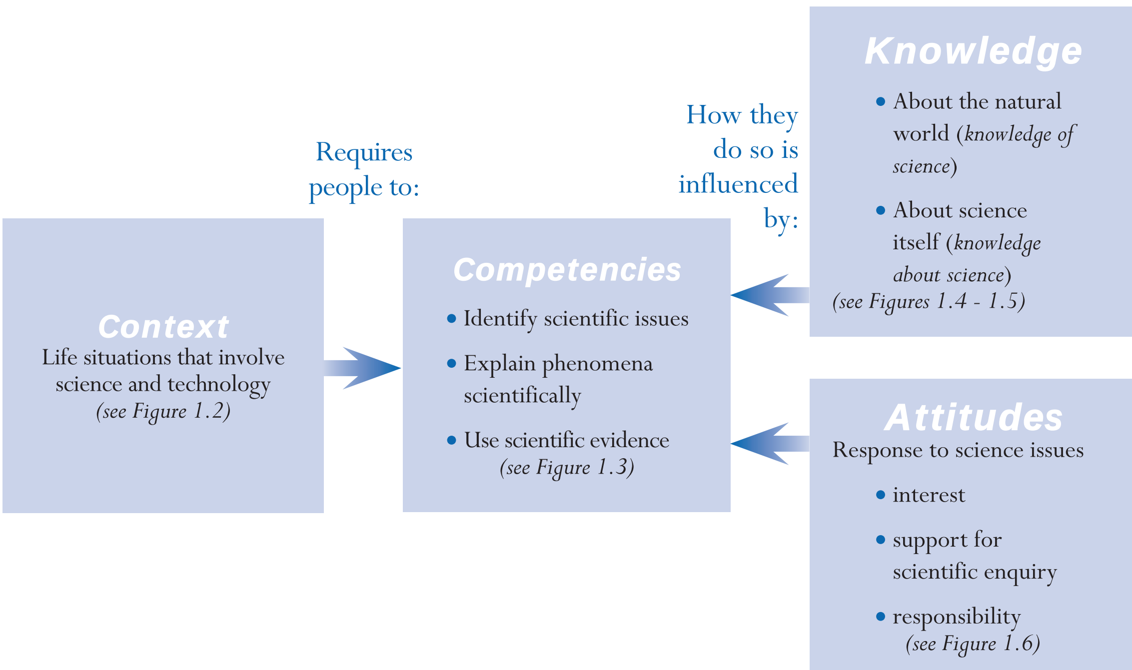


Figure 1.1 • **Framework for PISA 2006 science assessment**



The following sections restate and elaborate the interrelated aspects of *scientific literacy*. In highlighting these aspects, the PISA 2006 scientific literacy framework has ensured that the focus of the assessment is upon the outcomes of science education as a whole. Several questions, based on the PISA 2006 perspective of *scientific literacy* lie behind the organisation of this section of the framework. They are:

- What *contexts* would be appropriate for assessing 15-year-old students?
- What *competencies* might we reasonably expect 15-year-old students to demonstrate?
- What *knowledge* might we reasonably expect 15-year-old students to demonstrate?
- What *attitudes* might we reasonably expect 15-year-old students to demonstrate?

SITUATIONS AND CONTEXT

An important aspect of *scientific literacy* is engagement with science in a variety of situations. In dealing with scientific issues, the choice of methods and representations is often dependent on the situations in which the issues are presented.

The situation is the part of the student's world in which the tasks are placed. Assessment items are framed in situations of general life and not limited to life in school. In the PISA 2006 science assessment, the focus of the items is on situations relating to the self, family and peer groups (*personal*), to the community (*social*) and to life across the world (*global*). A further type of situation, appropriate to some topics, is the *historical* one, in which understanding of the advances in scientific knowledge can be assessed.

The context of an item is its specific setting within a situation. It includes all the detailed elements used to formulate the question.

PISA 2006 assesses important scientific knowledge relevant to the science education curricula of participating countries without being constrained to the common aspects of participants' national curricula. The assessment does this by requiring evidence of the successful use of scientific competencies in important situations reflecting the world and in accordance with PISA's focus on *scientific literacy*. This, in turn, involves the application of selected knowledge about the natural world, and about science itself, and evaluation of students' attitudes towards scientific matters.

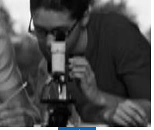
Figure 1.2 lists the applications of science, within *personal, social,* and *global* situations, primarily used as the contexts for assessment exercises. However, other situations (*e.g. technological, historical*) and areas of application are used. The applications were drawn from a wide variety of life situations and were generally consistent with the areas of application for *scientific literacy* in the 2000 and 2003 PISA frameworks. The areas of application are: "health", "natural resources", "the environment", "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 the development of public policy.

The PISA science assessment is not an assessment of contexts. It assesses competencies, knowledge and attitudes as these are presented or relate to contexts. In selecting the contexts, it is important to keep in mind that the purpose of the assessment is to assess scientific competencies, understandings, and attitudes that students have acquired by the end of the compulsory years of schooling.

The contexts used for assessment items are chosen in the light of relevance to students' interests and lives. Science items are developed keeping in mind linguistic and cultural differences in participating countries.

Figure 1.2 • **Contexts for the PISA 2006 science assessment**

	Personal (Self, family and peer groups)	**Social** (The community)	**Global** (Life across the world)
Health	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
Environment	Environmentally friendly behaviour, use and disposal of materials	Population distribution, disposal of waste, environmental impact, local weather	Biodiversity, ecological sustainability, control of pollution, production and loss of soil
Hazard	Natural and human-induced, decisions about housing	Rapid changes (earthquakes, severe weather), slow and progressive changes (coastal erosion, sedimentation), risk assessment	Climate change, impact of modern warfare
Frontiers of science and technology	Interest in science's explanations of natural phenomena, science-based hobbies, sport and leisure, music and personal technology	New materials, devices and processes, genetic modification, weapons technology, transport	Extinction of species, exploration of space, origin and structure of the universe



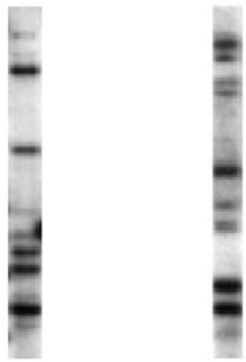
Science Example 1 is part of a unit titled *CATCHING THE KILLER*. The stimulus material is a newspaper article that establishes the context for the unit. The area of application is "Frontiers of science and technology" within a social setting.

Science Example 1: CATCHING THE KILLER

DNA TO FIND KILLER

Smithville, yesterday: *A man died from multiple stab wounds in Smithville today. Police say that there were signs of a struggle and that some of the blood found at the scene of the crime did not match the victim's blood. They believe that this blood came from the killer.*

To help find the killer, police scientists have prepared a DNA profile from the blood sample. When compared to DNA profiles of convicted criminals, kept on a computer database, no match was found.



Person A **Person B**

Photo of typical DNA profiles from two people. The bars are different fragments of each person's DNA. Each person has a different pattern of bars. Like fingerprints, these patterns can identify a person.

Police have now arrested a local man seen arguing with the victim earlier in the day. They have applied for permission to collect a sample of the suspect's DNA.

Sergeant Brown of the Smithville police said, "We just need to take a harmless scraping from the inside of the cheek. From this scraping scientists can extract DNA and form a DNA profile like the ones pictured."

Except for identical twins, there is only a 1 in 100 million chance that two people will have the same DNA profile.

Question 1: CATCHING THE KILLER

This newspaper article refers to the substance DNA. What is DNA?

A. A substance in cell membranes that stops the cell contents leaking out.

B. A molecule that contains the instructions to build our bodies.

C. A protein found in blood that helps carry oxygen to our tissues.

D. A hormone in blood that helps regulate glucose levels in body cells.

Question 2: CATCHING THE KILLER

Which one of the following questions *cannot* be answered by scientific evidence?

A. What was the medical or physiological cause of the victim's death?

B. Who was the victim thinking of when he died?

C. Is taking cheek scrapings a safe way to collect DNA samples?

D. Do identical twins have exactly the same DNA profile?

SCIENTIFIC COMPETENCIES

The PISA 2006 science assessment gives priority to the competencies listed in Figure 1.3, the ability to: identify scientifically-oriented issues; describe, explain or predict phenomena based on scientific knowledge; interpret evidence and conclusions; and use scientific evidence to make and communicate decisions. These competencies involve scientific knowledge – both knowledge of science and knowledge about science itself as a form of knowledge and an approach to enquiry.

Some cognitive processes have special meaning and relevance for *scientific literacy*. Among the cognitive processes that are implied in the scientific competencies are: inductive/deductive reasoning, critical and integrated thinking, transforming representations (*e.g.* data to tables, tables to graphs), constructing and communicating arguments and explanations based on data, thinking in terms of models, and using mathematics.

Justification for an emphasis on the scientific competencies of Figure 1.3 in PISA 2006 rests on the importance of these competencies for scientific investigation. They are grounded in logic, reasoning, and critical analysis. An elaboration of the scientific competencies follows.

Figure 1.3 • **PISA 2006 scientific competencies**

Identifying scientific issues

- Recognising issues that it is possible to investigate scientifically
- Identifying keywords to search for scientific information
- Recognising the key features of a scientific investigation

Explaining phenomena scientifically

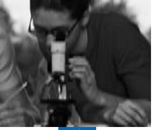
- Applying knowledge of science in a given situation
- Describing or interpreting phenomena scientifically and predicting changes
- Identifying appropriate descriptions, explanations, and predictions

Using scientific evidence

- Interpreting scientific evidence and making and communicating conclusions
- Identifying the assumptions, evidence and reasoning behind conclusions
- Reflecting on the societal implications of science and technological developments

Identifying scientific issues

It is important to be able to distinguish scientific issues and content from other forms of issues. Importantly, scientific issues must lend themselves to answers based on scientific evidence. The competency *identifying scientific issues* includes recognising questions that it would be possible to investigate scientifically in a given situation and identifying keywords to search for scientific



information on a given topic. It also includes recognising key features of a scientific investigation: for example, what things should be compared, what variables should be changed or controlled, what additional information is needed, or what action should be taken so that relevant data can be collected.

Identifying scientific issues requires students to possess knowledge about science itself, but may also draw, to varying degrees, on their knowledge of science. Question 2 of *CATCHING THE KILLER* (Science Example 1) requires students to identify a question that cannot be investigated scientifically. The item mainly assesses students' knowledge of what types of questions can be investigated scientifically (Knowledge about science, category: "Scientific enquiry"), but assumes a knowledge of science (category: "Living systems") that 15-year-old students could be expected to possess.

Explaining phenomena scientifically

Students demonstrate the competency *explaining phenomena scientifically* by applying appropriate knowledge of science in a given situation. The competency includes describing or interpreting phenomena and predicting changes, and may involve recognising or identifying appropriate descriptions, explanations, and predictions. Question 1 of *CATCHING THE KILLER* (Science Example 1) requires students to draw on their knowledge of science (category: "Living systems") to recognise the appropriate description of DNA.

Using scientific evidence

The competency *using scientific evidence* requires students to make sense of scientific findings as evidence for claims or conclusions. The required response can involve knowledge about science or knowledge of science or both. The question in *MALARIA* (Science Example 2) requires students to make conclusions based on the evidence presented about the life cycle of a mosquito. The item mainly assesses whether students can interpret a standard representation (model) of a life cycle – this is knowledge about science (category: "Scientific explanations" – see Figure 1.5).

Using scientific evidence includes accessing scientific information and producing arguments and conclusions based on scientific evidence (Kuhn, 1992; Osborne, Erduran, Simon and Monk, 2001). The competency may also involve: selecting from alternative conclusions in relation to evidence; giving reasons for or against a given conclusion in terms of the process by which the conclusion was derived from the data provided; and identifying the assumptions made in reaching a conclusion. Reflecting on the societal implications of scientific or technological developments is another aspect of this competency.

Students may be required to express their evidence and decisions, through their own words, diagrams or other representations as appropriate, to a specified audience. In short, students should be able to present clear and logical connections between evidence and conclusions or decisions.

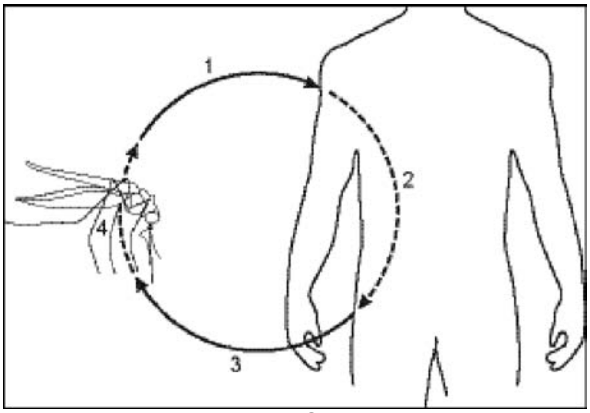
Science Example 2: MALARIA

Malaria is the cause of more than one million deaths every year. The fight against malaria is currently in crisis. Mosquitoes pass the malaria parasite from person to person. The malaria-carrying mosquito has become resistant to many pesticides. Also, medicines against the malaria parasite are getting less and less effective.

Life cycle of the malaria parasite

1
The malaria parasite is passed to a human through a female mosquito bite or sting.

4
The parasite also reproduces in the mosquito's body, but the mosquito is not affected.



2
The parasite reproduces in the human body and causes the first symptoms of the disease.

3
The parasite is transferred to another mosquito that bites or stings the infected person.

Question 1: MALARIA

Three methods of preventing the spread of malaria are given below.

Which of the stages (1, 2, 3 and 4) in the life cycle of a malaria parasite are *directly* affected by each method? *Circle the relevant stage(s) for each method (more than one stage may be affected by a single method).*

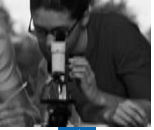
Method of preventing the spread of malaria	Stages in the life cycle of the parasite that are affected
Sleeping underneath a mosquito net.	1 2 3 4
Taking medicines against malaria.	1 2 3 4
Using pesticides against mosquitoes.	1 2 3 4

SCIENTIFIC KNOWLEDGE

As previously noted, scientific knowledge refers to both *knowledge of science* (knowledge about the natural world) and *knowledge about science* itself.

Knowledge of science

Given that only a sample of students' knowledge of science can be assessed in the PISA 2006 science assessment, it is important that clear criteria are used to guide the selection of knowledge that is assessed. Moreover, the objective of PISA is to describe the extent to which students can apply their knowledge in contexts of relevance to their lives. Accordingly, the assessed knowledge



will be selected from the major fields of physics, chemistry, biology, Earth and space science, and technology[3] according to the following criteria:

- Relevance to real-life situations: scientific knowledge differs in the degree to which it is useful in the life of individuals
- Knowledge selected represents important scientific concepts and thus has enduring utility
- Knowledge selected is appropriate to the developmental level of 15-year-old students

Figure 1.4 shows the *knowledge of science* categories and examples of content selected by applying these criteria. This knowledge is required for understanding the natural world and for making sense

Figure 1.4 • **PISA 2006 categories of *knowledge of science***

Physical systems

- 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.* reactions, energy transfer, acids/bases)
- Motions and forces (*e.g.* velocity, friction)
- Energy and its transformation (*e.g.* conservation, dissipation, chemical reactions)
- Interactions of energy and matter (*e.g.* light and radio waves, sound and seismic waves)

Living systems

- Cells (*e.g.* structures and function, DNA, plant and animal)
- Humans (*e.g.* health, nutrition, subsystems [i.e. digestion, respiration, circulation, excretion, and their relationship], disease, reproduction)
- 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

- 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)

Technology systems

- Role of science-based technology (*e.g.* solve problems, help humans meet needs and wants, design and conduct investigations)
- Relationships between science and technology (*e.g.* technologies contribute to scientific advancement)
- Concepts (*e.g.* optimisation, trade-offs, cost, risk, benefit)
- Important principles (*e.g.* criteria, constraints, innovation, invention, problem solving)

of experiences in *personal, social* and *global* contexts. For these reasons the framework uses the term "systems" instead of "sciences" in the descriptors of the major fields. The intention is to convey the idea that citizens have to understand concepts from the physical and life sciences, Earth and space science, and technology, in several different contexts.

The examples listed in Figure 1.4 convey the meanings of the categories; there is no attempt to list comprehensively all the knowledge that could be related to each of the knowledge of science categories. Question 1 of *CATCHING THE KILLER* (Science Example 1) assesses students' *knowledge of science* in the category "Living systems".

Knowledge about science

Figure 1.5 displays the categories and examples of content for *knowledge about science*. The first category, "Scientific enquiry," centres on enquiry as the central process of science and the various components of that process. The second category, closely related to enquiry, is "Scientific explanations". Scientific explanations are the results of scientific enquiry. One can think of enquiry as the means of science (how scientists get data) and explanations as the goals of science (how scientists use data). The examples listed in Figure 1.5 convey the general meanings of the categories; there is no attempt to list comprehensively all the knowledge that could be related to each category.

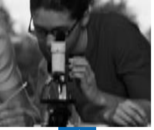
Figure 1.5 • **PISA 2006 categories of *knowledge about science***

Scientific enquiry

- Origin (*e.g.* curiosity, scientific questions)
- Purpose (*e.g.* to produce evidence that helps answer scientific questions, current ideas/models/theories guide enquiries)
- Experiments (*e.g.* different questions suggest different scientific investigations, design)
- Data type (*e.g.* quantitative [measurements], qualitative [observations])
- Measurement (*e.g.* inherent uncertainty, replicability, variation, accuracy/precision in equipment and procedures)
- Characteristics of results (*e.g.* empirical, tentative, testable, falsifiable, self-correcting)

Scientific explanations

- Types (*e.g.* hypothesis, theory, model, law)
- Formation (*e.g.* data representation, role of extant knowledge and new evidence, creativity and imagination, logic)
- Rules (*e.g.* must be logically consistent; based on evidence, historical and current knowledge)
- Outcomes (*e.g.* produce new knowledge, new methods, new technologies; lead to new questions and investigations)



Science Example 3 is part of a unit titled *SCHOOL MILK STUDY*, with a historical setting and health as the area of application. Both questions assess students' knowledge about science, in the category "Scientific enquiry". Question 1 requires students to identify the possible purposes of the study (competency: "Identifying scientific issues"). The competency classification of Question 2 is also "Identifying scientific issues" (rather than "Using scientific evidence") since the most obvious assumption (that the three groups of students were not significantly different in any relevant way) relates to the design of the study.

Science Example 3: SCHOOL MILK STUDY

In 1930, a large-scale study was carried out in the schools in a region of Scotland. For four months, some students received free milk and some did not. The head teachers in each school chose which of their students received milk. Here is what happened:

- *5 000 school children received an amount of unpasteurised milk each school day*
- *Another 5 000 school children received the same amount of pasteurised milk*
- *10 000 school children did not receive any milk at all*

All 20 000 children were weighed and had their heights measured at the beginning and the end of the study.

Question 1: SCHOOL MILK STUDY

Is it likely that the following questions were research questions for the study?

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

Is it likely that this was a research question for the study?	Yes or No?
What has to be done to pasteurise milk?	Yes / No
What effect does the drinking of additional milk have on school children?	Yes / No
What effect does milk pasteurisation have on school children's growth?	Yes / No
What effect does living in different regions of Scotland have on school children's health?	Yes / No

---

Question 2: SCHOOL MILK STUDY

On average, the children who received milk during the study gained more in height and weight than the children who did not receive milk.

One possible conclusion from the study, therefore, is that school children who drink a lot of milk grow faster than those who do not drink a lot of milk.

To have confidence in this conclusion, indicate one assumption that needs to be made about these two groups of students in the study.

ATTITUDES TOWARDS SCIENCE

Peoples' attitudes play a significant role in their interest, attention, and response to science and technology in general and to issues that affect them in particular. One goal of science education is for students to develop attitudes that make them likely to attend to scientific issues and subsequently to acquire and apply scientific and technological knowledge for personal, social, and global benefit.

The PISA 2006 assessment of science takes an innovative approach to assessing student attitudes. Not only does it ask them about what they think about science in the student questionnaire, but it also asks them, in the course of the science part of the assessment, what their attitudes are towards the issues that they are being tested on.

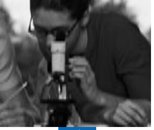
The survey's attention to attitudes towards science is based on the belief that a person's *scientific literacy* includes certain attitudes, beliefs, motivational orientations, sense of self-efficacy, values, and ultimate actions. The inclusion of attitudes and the specific areas selected for PISA 2006 is supported by and builds upon Klopfer's (1976) structure for the affective domain in science education as well as reviews of attitudinal research (for example, Gardner, 1975, 1984; Gauld and Hukins, 1980; Blosser, 1984; Laforgia, 1988; Schibeci, 1984).

The PISA 2006 science assessment evaluated students' attitudes in three areas: *interest in science*, *support for scientific enquiry* and *responsibility towards resources and environments* (see Figure 1.6). These areas were selected because they provide an international portrait of students' general appreciation of science, their specific scientific attitudes and values, and their responsibility towards selected science-related issues that have national and international ramifications. This was not an assessment of students' attitudes toward school science programs or teachers. The results may provide information about the emerging problem of declining enrolments for science studies among young people.

Interest in science was selected because of its established relationships with achievement, course selection, career choice, and lifelong learning. The relationship between (individual) interest in science and achievement has been the subject of research for more than 40 years although there is still debate about the causal link (see, for example, Baumert and Köller, 1998; Osborne, Simon & Collins, 2003). The PISA 2006 science assessment addressed students' interest in science through knowledge about their engagement in science-related social issues, their willingness to acquire scientific knowledge and skills, and their consideration of science-related careers.

Support for scientific enquiry is widely regarded as a fundamental objective of science education and as such warrants assessing. It is a similar construct to "adoption of scientific attitudes" as identified by Klopfer (1971). Appreciation of and support for scientific enquiry implies that students value scientific ways of gathering evidence, thinking creatively, reasoning rationally, responding critically, and communicating conclusions as they confront life situations related to science. Aspects of this area in PISA 2006 include the use of evidence (knowledge) in making decisions, and the appreciation for logic and rationality in formulating conclusions.

Responsibility towards resources and environments is of international concern, as well as being of economic relevance. Attitudes in this area have been the subject of extensive research since the 1970s (see, for example, Bogner and Wiseman, 1999; Eagles & Demare, 1999; Weaver, 2002; Rickinson, 2001). In December 2002, the United Nations approved resolution 57/254 declaring



the ten-year period beginning on 1 January 2005 to be the United Nations Decade of Education for Sustainable Development (UNESCO, 2003). The International Implementation Scheme (UNESCO, September 2005) identifies environment as one of the three spheres of sustainability (along with society (including culture) and economy) that should be included in all education for sustainable development programmes.

PISA 2006 gathers data about such student attitudes both by posing questions in the student questionnaire and in contextualised test items – that is questions about attitudes towards issues posed immediately after test questions related to these issues (see Box 1.2). The student questionnaire collects information about students' attitudes in all three areas: *interest in science*, *support for scientific enquiry* and *responsibility towards resources and environments,* in a non-contextualised manner. Additional data concerning students' engagement in science (*e.g.* self-efficacy, enjoyment of science and frequency of out of school scientific activities) were also collected via the student questionnaire, as were students' views on the value of science for their own lives (*e.g.* further education and career choices) and for society (*e.g.* social and economic benefits).

Contextualised items are used in relation to interest in learning about science and student support for scientific enquiry. Contextualised items add value to the assessment in that they provide data on whether students' attitudes differ when assessed in or out of context, whether they vary between contexts, and whether they correlate with performance at the unit level. One aspect of students' *interest in science* (namely, their interest in learning about science) and of students' *support for scientific enquiry* was assessed in the test using embedded items that targeted *personal, social,* and *global* issues.

The results of PISA 2006 will provide important information for educational policy makers in the participating countries. The combined richness of the data obtained through both the student questionnaire and the embedded attitudinal items should generate new knowledge about students' predispositions towards scientifically literate behaviours. Further, since the literature contains conflicting reports on the correlation between attitudes and performance in science, it remains to be seen how student attitudinal data (concerning students' *interest in science*, *support for scientific enquiry* and *responsibility towards resources and environments*), collected via the test and the questionnaire, correlates with student performance. Other data obtained from the student questionnaire, such as students' engagement in science and science-related behaviours, also will be reported and linked with student performance.

ASSESSING SCIENTIFIC LITERACY

Test characteristics

In accordance with the PISA definition of *scientific literacy,* test questions (items) require the use of the scientific competencies (see Figure 1.3) within a context (see Figure 1.2). This involves the application of scientific knowledge (see Figures 1.4 and 1.5) and reflects aspects of the respondents' attitudes towards scientific matters (see Figure 1.6).

Figure 1.7 is a variation of Figure 1.1 that presents the basic components of the PISA framework for the 2006 scientific literacy assessment in a way that can be used to relate the framework with the structure and the content of assessment units. Figure 1.7 may be used both synthetically as a tool to plan assessment exercises, and analytically as a tool to study the results of standard assessment exercises. As a starting point to construct assessment units, we could consider the contexts that

Figure 1.6 • **PISA 2006 areas for assessment of attitudes**

Interest in science

- Indicate curiosity in science and science-related issues and endeavours
- Demonstrate willingness to acquire additional scientific knowledge and skills, using a variety of resources and methods
- Demonstrate willingness to seek information and have an ongoing interest in science, including consideration of science-related careers

Support for scientific enquiry

- Acknowledge the importance of considering different scientific perspectives and arguments
- Support the use of factual information and rational explanations
- Express the need for logical and careful processes in drawing conclusions

Responsibility towards resources and environments

- Show a sense of personal responsibility for maintaining a sustainable environment
- Demonstrate awareness of the environmental consequences of individual actions
- Demonstrate willingness to take action to maintain natural resources

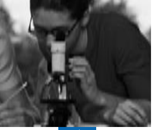
would serve as stimulus material, the competencies required to respond to the questions or issues, or the knowledge and attitudes central to the exercise.

A test unit comprises specific stimulus material, which may be a brief written passage, or text accompanying a table, chart, graph, or diagram, plus items which are a set of independently scored questions of various types, as illustrated by the three examples (*CATCHING THE KILLER, MALARIA* and *SCHOOL MILK STUDY*) already discussed and the additional examples included in the Annex A.

The reason PISA employs this unit structure is to facilitate the employment of contexts that are as realistic as possible, and that reflect the complexity of real 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 relating to each question. However, the need to make each scored 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 2006 test units incorporate up to four cognitive items which assess students' scientific competencies. Each item involves the predominant use of one of the scientific competencies and requires mainly knowledge of science or knowledge about science. In most cases, more than one competency and more than one knowledge category was assessed (by different items) within a unit.

Four types of items were used to assess the competencies and scientific knowledge identified in the framework. About one-third of the items were (simple) multiple-choice items, which required the

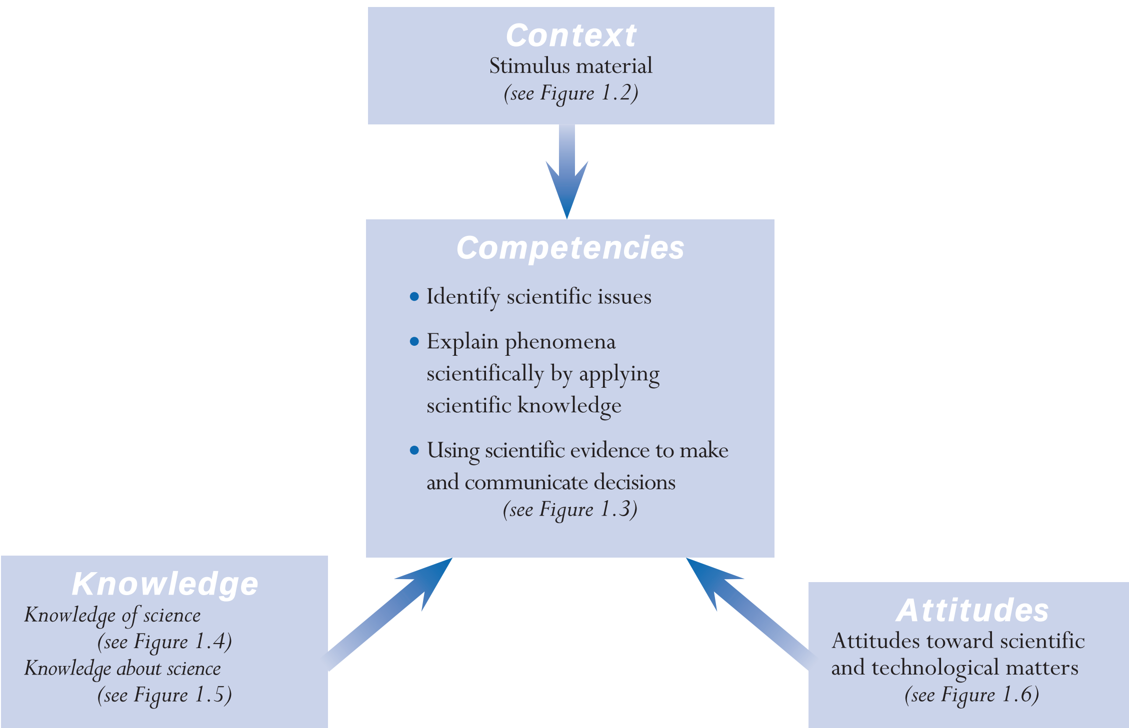


selection of a single response from four options. A further third either required short constructed responses, like Question 1 of *MALARIA* (Science Example 2), or complex multiple-choice items. Question 1 of *SCHOOL MILK STUDY* (Science Example 3), which requires students to respond to a series of related "Yes/No" questions, is a typical complex multiple-choice item. The remaining one-third of the items were open-constructed response items, like Question 2 in *SCHOOL MILK STUDY* (Science Example 3), that required a relatively extended written or drawn response from a student.

Multiple-choice and short-constructed response items can be used to validly assess most of the cognitive processes involved in the three scientific competencies and open-response items provide the opportunity to assess the ability to communicate.

Although the majority of the items are dichotomously scored (that is, credit or no credit), some of the complex multiple-choice and open-response items will involve partial credit scoring, which give students credit for getting part of the question correct, but not the whole question. For each partial credit item, a detailed coding guide that allows for "Full credit", "Partial credit" and "No credit" is provided. The categories "Full credit", "Partial credit" and "No credit" divide students' responses into three groups in terms of the extent to which the students demonstrate ability to answer the question. A "Full credit" response, although not necessarily absolutely scientifically correct, requires a student to exhibit a level of understanding of the topic appropriate for a scientifically literate 15-year-old student. Less sophisticated or correct responses may qualify for "Partial credit", with completely incorrect, irrelevant or missing responses being assigned "No credit". Question 1 of *MALARIA* (Science Example 2) is a partial credit item and its scoring scheme (coding guide) is shown in Science Example 4.

Figure 1.7 • **A tool for constructing and analysing assessment units and items**



Science Example 4: MALARIA (Question 1 scoring)

Full Credit

Code 2: All three correct: [1 and 3]; [2]; and [1, 3 and 4] in that order

Partial Credit

Code 1: Two of the three rows correct

OR

One (or more) correct, ***but none wrong****, in each row*

No Credit

Code 0: Other responses

Code 9: Missing

Most of the new units included in the PISA 2006 science test also contain an item that assesses students' *interest in learning about science* or an item that assesses *support for scientific enquiry* or both types of items. Question 3 of the unit *CATCHING THE KILLER*, included below as Science Example 5, is an example of this. This item requires students to indicate their level of interest in three tasks to assess their interest in learning more about the application of science to solving crime. A unipolar response format ("High interest", "Medium interest", "Low interest", "No interest"), rather than the conventional bipolar one ("Strongly agree", "Agree", "Disagree", "Strongly disagree"), is used in this example to reduce the influence of social desirability on responses.

Science Example 5: CATCHING THE KILLER (Attitudinal item)

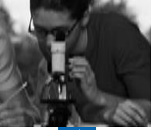
Question 3: CATCHING THE KILLER

How much interest do you have in the following information?

Tick only one box in each row.

		High Interest	Medium Interest	Low Interest	No Interest
a)	Knowing more about the use of DNA in solving crime.	☐1	☐2	☐3	☐4
b)	Learning more about how DNA profiling works.	☐1	☐2	☐3	☐4
c)	Understanding better how crime can be solved using science.	☐1	☐2	☐3	☐4

On the actual test form given to students, attitudinal items are distinctively formatted in a shaded box to remind students that, for each statement, they should tick the box that indicates their own opinion about the statement. In addition, the general directions at the start of each booklet include the following instruction:



Some of the questions are about your attitude or opinion regarding certain issues. These questions are set out differently from the others – they appear inside a shaded box. THERE IS NO CORRECT ANSWER to these questions and they will not count in your test score, but it is important that you answer them truthfully.

The need for students to have a degree of *reading literacy* in order to understand and answer written questions on *scientific literacy* raises an issue of the level of *reading literacy* required. Stimulus material and questions used language that is as clear, simple and brief as possible while still conveying the appropriate meaning. The number of concepts introduced per paragraph was limited and questions that predominantly assess *reading literacy* or *mathematical literacy* were avoided.

Science assessment structure

It is important that the test includes an appropriate balance of items assessing the various components of the scientific literacy framework. Figure 1.8 shows the desired balance of items relating to the knowledge of science versus knowledge about science. The balance is expressed in terms of the percentage of total score points allocated to each category. Figure 1.8 also shows the desired distribution of score points among the various knowledge of science and knowledge about science categories.

Figure 1.8 • **Desired distribution of score points for knowledge**

Knowledge of science	**Per cent of score points**
Physical systems	15-20
Living systems	20-25
Earth and space systems	10-25
Technological systems	5-10
Subtotal	*60-65*
Knowledge about science	
Scientific enquiry	15-20
Scientific explanation	15-20
Subtotal	*35-40*
Total	***100***

The desired balance for scientific competencies is given in Figure 1.9.

Figure 1.9 • **Desired distribution of score points for scientific competencies**

Scientific competencies	**Per cent of score points**
Identifying scientific issues	25-30
Explaining phenomenas scientifically	35-40
Using scientific evidence	35-40
Total	***100***

Item contexts are spread across personal, social and global settings roughly in the ratio 1:2:1. A wide selection of areas of application were used for units, subject to satisfying as far as possible the various constraints imposed by the previous two paragraphs.

About 60% of the units contain one or two attitudinal items that assess students' *interest in learning about science* or their *support for scientific enquiry*. Responding to these items occupies about 11% of the total test time. To facilitate comparability of performance over time, link items included from the two previous PISA science assessments did not contain attitudinal items.

Reporting scales

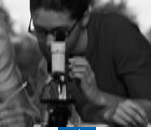
To meet the aims of PISA, the development of scales of student achievement is essential. The process of arriving at a scale has to be iterative. Initial descriptions, based on the results of the trials and the PISA 2000 and 2003 surveys – and informed by past experience of assessing science achievement and findings from research into learning and cognitive development in science – are likely to be modified as more data are accumulated in this and future surveys.

The construction of scales is facilitated by the inclusion of items which have a wide range of difficulties. Factors that determine difficulty in items assessing science achievement include the:

- General complexity of the context
- Level of familiarity with the scientific ideas, processes and terminology involved
- Length of the train of logic required to respond to a question – that is, the number of steps needed to arrive at an adequate response and the level of dependence of each step on the previous one
- Degree to which abstract scientific ideas or concepts are required in forming a response
- Level of reasoning, insight and generalisation involved in forming judgements, conclusions, and explanations

For PISA 2000, when science was a minor domain and thus having limited testing time, students' science achievement was reported in terms of a proficiency scale with a mean of 500 and a standard deviation of 100. Although no proficiency levels were identified, it was possible to describe what processes (*i.e.* scientific competencies) students can perform at three points in this scale (OECD, 2001):

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



- Towards the lower end of the scale (around 400 points), students are able to recall simple factual scientific knowledge (*e.g.* names, facts, terminology, simple rules); and to use common scientific knowledge in drawing or evaluating conclusions.

For PISA 2003, the reporting of science results followed a similar format to that of 2000 (OECD, 2004). However, with science as the major domain for the PISA 2006 assessment, the increased testing time available should enable the construction of separate scales based on either the scientific competencies or the two knowledge components.

Proficiency in science in PISA 2000 and 2003 was described on a scale in terms of scientific competencies as shown in the Figure 1.3. By examining the descriptions we can derive the skeleton of each PISA 2006 competency scale. For example, the skeleton scale shown in Figure 1.10 can be derived from the competency "Using scientific evidence".

Figure 1.10 • **Example of a competency-based reporting scale**

High	Can compare data in order to evaluate alternative viewpoints or differing perspectives; can communicate scientific arguments and/or descriptions in detail and with precision.
	Able to select relevant information from competing data or chains of reasoning in drawing or evaluating conclusions.
Low	Able to use common scientific knowledge in drawing or evaluating conclusions.

Alternatively, it should be possible to report separate scales for the two knowledge components, knowledge of science and knowledge about science. The competencies would then be central to describing the proficiency levels for these two knowledge scales. Decisions about the actual scales to be reported, and the number of proficiency levels to be identified, will be made following analysis of the PISA 2006 assessment data.

It should also be possible to prepare reliable scales for *interest in science* and *support for scientific enquiry* using the data obtained from the embedded attitudinal items and the student questionnaire. A scale for *responsibility towards resources and environments* will be constructed from data obtained from the student questionnaire.

The scores on attitude items will not be included in an index (or overall score) of *scientific literacy;* rather, they will form a component of a profile of student s*cientific literacy*.

SUMMARY

Science was the major testing domain for the first time in PISA 2006. The definition of *scientific literacy* has been elaborated and expanded from that used in PISA 2000 and 2003. A major innovation is to include students' attitudinal responses towards scientific issues, not just in an accompanying questionnaire but in additional questions about attitudes to scientific issues juxtaposed with test questions relating to the same issues. In addition, there is an increased emphasis on students' understanding of the nature and methodology of science itself (their knowledge about science), and of the role of science-based technology.

The PISA 2006 definition of *scientific literacy* has its origin in the consideration of what 15-year-old students should know, value and be able to do as preparedness for life in modern society. Central to the definition, and the assessment of *scientific literacy,* are the competencies that are characteristic of science and scientific enquiry. The ability of students to perform these competencies depends on their scientific knowledge, both knowledge of the natural world and knowledge about science itself, and their attitudes towards science-related issues.

This framework describes and illustrates the scientific competencies, knowledge and attitudes being assessed in PISA 2006 (see Figure 1.11), and the contexts for test items. Test items were grouped into units with each unit beginning with stimulus material that establishes the context for its items. A combination of item types was used and some items involved partial credit scoring. Attitudinal items were embedded in over half of the units and occupied about 11% of testing time.

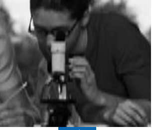
Figure 1.11 • **Major components of the PISA 2006 assessment of scientific literacy**

Competencies	Knowledge	Attitudes
Identifying scientific issues	Knowledge of science:	Interest in science[1]
Explaining scientific phenomena	Physical systems	Support for scientific enquiry
Using scientific eveidence	Living systems	Responsibility towards resources and environment[2]
	Earth and space systems	
	Knowledge about science:	
	Scientific enquiry	
	Scientific explanations	

1. Embedded items assess "Interest in learning about science"
2. Not assessed with embedded items

The ratio of items assessing students' knowledge of science, to items assessing their knowledge about science, was about 3:2, while each of the three scientific competencies were assessed by at least 25% of the items. This should enable separate scales, with described proficiency levels, to be constructed for each of the competencies, or for the two types of knowledge. Scales should also be able to be constructed for the attitudes that were assessed with embedded items.

Further examples to illustrate the PISA science assessment framework are included in Annex A.



Notes

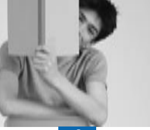
1. Throughout this framework, "natural world" includes the changes made by human activity, including the "material world" designed and shaped by technologies.

2. The PISA science concept of "literacy" can be compared to the DeSeCo (OECD, 2003b) definition of "competency" in that both involve attitudes and values, as well as knowledge and skills.

3. Knowledge of the design or internal working of technology artefacts (*e.g.* aeroplanes, engines, computers) will not be assumed.

2

Reading Literacy



DEFINITION OF THE DOMAIN

Definitions of reading and *reading literacy* have changed over time in parallel with changes in society, the economy and culture. The concept of learning, and particularly the concept of lifelong learning, has expanded perceptions of *reading literacy* and the demands made on it. Literacy is no longer considered an ability only acquired in childhood during the early years of schooling. Instead, it is viewed as an expanding set of knowledge, skills and strategies which individuals build on throughout life in various situations and through interaction with their peers and with the larger communities in which they participate.

Through a consensus-building process involving the reading experts selected by the participating countries and the PISA advisory groups, the following definition of *reading literacy* was adopted for the survey:

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

This definition goes beyond the notion of *reading literacy* as decoding and literal comprehension: it implies that *reading literacy* involves understanding, using and reflecting on written information for a variety of purposes. It thus takes into account the active and interactive role of the reader in gaining meaning from written texts. The definition also recognises the full scope of situations in which *reading literacy* plays a role for young adults, from private to public, from school to work, from active citizenship to lifelong learning. It spells out the idea that literacy enables the fulfilment of individual aspirations – from defined aspirations such as gaining an educational qualification or obtaining a job to those less immediate goals which enrich and extend one's personal life. Literacy also provides the reader with a set of linguistic tools that are increasingly important for meeting the demands of modern societies with their formal institutions, large bureaucracies and complex legal systems.

Readers respond to a given text in a variety of ways as they seek to use and understand what they are reading. This dynamic process involves many factors, some of which can be manipulated in large-scale assessments such as PISA. These include the reading situation, the structure of the text itself and the characteristics of the questions that are asked about the text (the test rubric). All of these factors are regarded as important components of the reading process and were considered in the creation of the items used in the assessment.

In order to use text format, characteristics of the items and situations in constructing the assessment tasks, and later in interpreting the results, the range for each of these factors had to be specified. This allowed for the categorisation of each task so that the weighting of each component could be taken into account in the final assembly of the survey.

TEXT FORMAT

At the heart of the PISA assessment is a distinction between continuous and non-continuous texts.

- *Continuous texts* are typically composed of sentences that are, in turn, organised into paragraphs. These may fit into even larger structures such as sections, chapters and books. The primary classification of continuous texts is by rhetorical purpose, that is, by text type.

- *Non-continuous texts* (or documents, as they are known in some approaches) can be categorised in two ways. One is the formal structure approach used in the work of Kirsch and Mosenthal (1989-1991). Their work classifies non-continuous texts by the way underlying lists are put. This approach is useful for understanding the similarities and differences between types of non-continuous texts. The other method of classification is by everyday descriptions of the formats of these texts. This second approach is used in classifying non-continuous texts in PISA.

Continuous texts

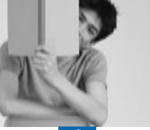
Text types are standard ways of organising continuous texts by content and author's purpose.

- *Narration* is the type of text in which the information refers to properties of objects in time. Narrative texts typically provide answers to "when", or "in what sequence" questions.
- *Exposition* is a type of text in which the information is presented as composite concepts or mental constructs, or elements into which concepts or mental constructs can be analysed. The text provides an explanation of how the component elements interrelate in a meaningful whole and often answers "how" questions.
- *Description* is a type of text in which the information refers to properties of objects in space. Descriptive texts typically provide an answer to "what" questions.
- *Argumentation* is a type of text that presents propositions as to the relationship between concepts, or other propositions. Argumentative texts often answer "why" questions. Another important sub-classification of argumentative texts is persuasive texts.
- *Instruction* (sometimes called injunction) is the type of text that provides directions on what to do and includes procedures, rules, regulations and statutes specifying certain behaviours.
- *Documents or records* are texts designed to standardise and conserve information. They can be characterised by highly formalised textual and formatting features.
- *Hypertext* is a set of text slots linked together in such a way that the units can be read in different sequences, allowing readers to follow various routes to the information.

Non-continuous texts

Non-continuous texts are organised differently from continuous texts and so require different kinds of reading approaches. Classifying non-continuous texts by their format, as shown below, provides a familiar means of discussing what types of non-continuous texts may be included in the assessment.

- *Charts and graphs* are iconic representations of data. They are used for the purposes of scientific argumentation, and also in journals and newspapers to display numerical and tabular public information in a visual format.
- *Tables* are row and column matrices. Typically, all the entries in each column and each row share properties and thus the column and row labels are part of the information structure of the text. Common tables include schedules, spreadsheets, order forms and indexes.
- *Diagrams* often accompany technical descriptions (*e.g.* demonstrating parts of a household appliance), expository texts and instructive texts (*e.g.* illustrating how to assemble a household appliance). It is often useful to distinguish procedural (how to) from process (how something works) diagrams.



- *Maps* are non-continuous texts that indicate the geographical relationships between places. There is a variety of types of maps. Road maps mark the distance and routes between identified places. Thematic maps indicate the relationships between locations and social or physical features.
- *Forms* are structured and formatted texts which request the reader to respond to specific questions in specified ways. Forms are used by many organisations to collect data. They often contain structured or pre-coded answer formats. Typical examples are tax forms, immigration forms, visa forms, application forms, statistical questionnaires, etc.
- *Information sheets* differ from forms in that they provide, rather than request, information. They summarise information in a structured way and in such a format that the reader can easily and quickly locate specific pieces of information. Information sheets may contain various text forms as well as lists, tables, figures and sophisticated text-based graphics (headings, fonts, indentation, borders, etc.) to summarise and highlight information. Timetables, price lists, catalogues and programmes are examples of this type of non-continuous text.
- *Calls and advertisements* are documents designed to invite the reader to do something, *e.g.* to buy goods or services, attend gatherings or meetings, elect a person to a public office, etc. The purpose of these documents is to persuade the reader. They offer something and request both attention and action. Advertisements, invitations, summonses, warnings and notices are examples of this document format.
- *Vouchers* testify that their owner is entitled to certain services. The information that they contain must be sufficient to show whether the voucher is valid or not. Typical examples are tickets, invoices, etc.
- *Certificates* are written acknowledgements of the validity of an agreement or a contract. They are formalised in content rather than format. They require the signature of one or more persons authorised and competent to bear testimony of the truth of the given statement. Warranties, school certificates, diplomas, contracts, etc. are documents that have these properties.

The distribution and variety of texts that students are asked to read for PISA are important characteristics of the assessment. Figure 2.1 shows the distributions of tasks for continuous and non-continuous texts in PISA 2000 (reading as major domain) and in PISA 2003 and 2006 (reading as minor domain). It can be readily seen that in 2000, 2003 and 2006 cycles continuous texts represent two-thirds of the tasks or items contained in the assessment. Within this category, in the three cycles, the largest percentage comes from expository texts.

CHARACTERISTICS OF THE ITEMS

Three sets of variables are used to describe the characteristics of the items: the processes (aspects), which set out the task for the student; item types, which set out the ways in which students are asked to demonstrate their proficiency at the task; and rules for coding, which specify how students' answers are to be evaluated. Each of these will be discussed in turn, though the first requires considerably more attention.

Five processes (aspects)

In an effort to simulate authentic reading situations, the PISA reading assessment measures the following five processes associated with achieving a full understanding of a text, whether the text

Figure 2.1 • Distribution of reading literacy tasks, by text format and type

Legend: grey = Reading as a major domain (PISA 2000); blue = Reading as a minor domain (PISA 2003 and 2006)

Text format and type	Percentage of tasks by text format and type (%) – Reading as a major domain (PISA 2000)	Percentage of tasks by text format and type (%) – Reading as a minor domain (PISA 2003 and 2006)	Percentage of tasks by text format and type, based on the whole test (%) – Reading as a major domain (PISA 2000)	Percentage of tasks by text format and type, based on the whole test (%) – Reading as a minor domain (PISA 2003 and 2006)
Continuous				
Narrative	21	17	14	11
Expository	36	67	24	43
Descriptive	14	17	9	11
Argumentative and persuasive	20	-	13	-
Injunctive	10	-	7	-
Total[1]	**100**	**100**	**68**	**64**
Non-continuous				
Charts and graphs	37	20	12	7
Tables	29	40	9	14
Diagrams	12	-	4	-
Maps	10	10	3	4
Forms	10	30	3	11
Advertisements	2	-	1	-
Total[1]	**100**	**100**	**34**	**37**

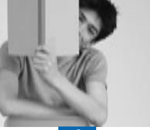
1. Data may not always add up to the totals indicated because of roundings.

is continuous or non-continuous. Students are expected to demonstrate their proficiency in all of these processes:

- *Retrieving information*
- *Forming a broad general understanding*
- *Developing an interpretation*
- *Reflecting on and evaluating the content of a text*
- *Reflecting on and evaluating the form of a text*

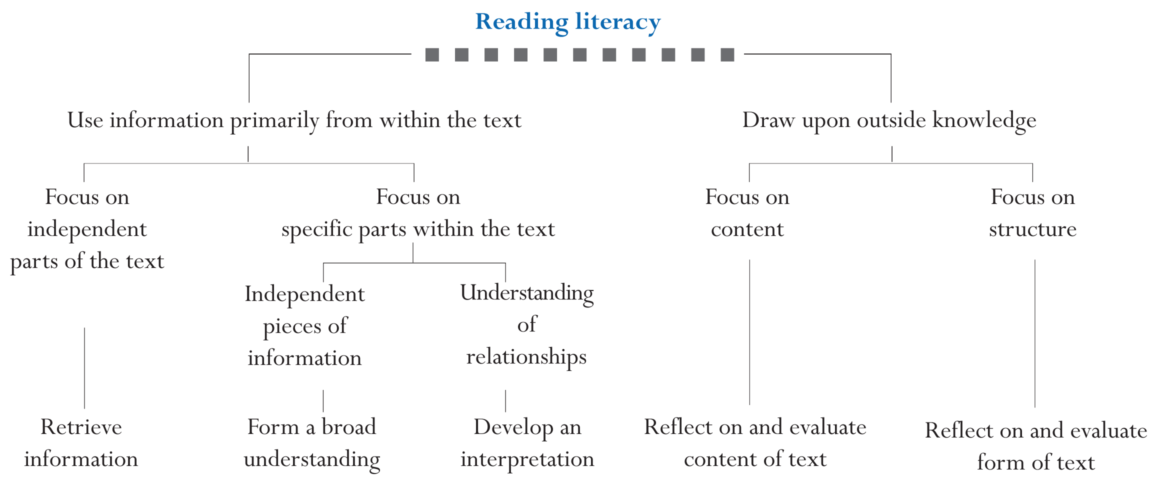
The full understanding of texts involves all of these processes. It is expected that all readers, irrespective of their overall proficiency, will be able to demonstrate some level of competency in each of them (Langer, 1995). While there is an interrelationship between the five aspects – each may require many of the same underlying skills – successfully accomplishing one may not ensure successful completion of any other. Some view them as being in the repertoire of each reader at every developmental level rather than forming a sequential hierarchy or set of skills.

Figure 2.2 identifies the key distinguishing characteristics of the five processes of reading measured in PISA. While this figure necessarily oversimplifies each process, it provides a useful scheme for organising and remembering the relationships between them. As depicted in this figure, the five processes can be distinguished in terms of four characteristics. The first deals with the extent to which the reader is expected to use information primarily from within the text or to draw also upon outside knowledge. A second characteristic involves the extent to which the reader is asked to focus on independent parts of the text or on the relationships within the information contained



in the text. Sometimes readers are expected to retrieve independent pieces of information while at other times they are asked to demonstrate their understanding of the relationships between parts of the text. Focusing on either the whole text or on relationships between parts of the text is the third distinguishing characteristic. The fourth characteristic relates to whether the reader is asked to deal with the content or substance of the text rather than its form or structure. The five processes of reading are represented in the last line of Figure 2.2 at the ends of the various branches. By starting at the top of the figure and following each branch one can see which characteristics are associated with each process.

Figure 2.2 • **Characteristics distinguishing the five processes (aspects) of *reading literacy***



The following discussion attempts to define each process operationally and to associate it with particular kinds of items. Although each process is discussed in terms of a single text, each can also apply to multiple texts when these are presented together as a unit within the test. The description of each process has two parts. The first provides a general overview of the process, while the second describes particular ways in which the process might be assessed.

Retrieving information

In the course of daily life, readers often need a particular piece of information: a telephone number or the departure time for a bus or train. They may want to find a particular fact to support or refute a claim someone has made. In situations such as these, readers are interested in retrieving isolated pieces of information. To do so, readers must scan, search for, locate and select relevant information. The processing involved is most frequently at the sentence level, though in some cases the information may be in two or more sentences or in different paragraphs.

In assessment tasks that call for retrieving information, students must match information given in the question with either identically worded or synonymous information in the text and use this to find the new information called for. In these tasks, *retrieving information* is based on the text itself and on explicit information included in it. Retrieving tasks require the student to find information based on requirements or features specified in questions. The student has to detect or identify one

or more essential elements of a question: characters, place/time, setting, etc. and then to search for a match that may be literal or synonymous.

Retrieving tasks can involve various degrees of ambiguity. For example, the student may be required to select explicit information, such as an indication of time or place in a text or table. A more difficult version of this same type of task might involve finding synonymous information. This sometimes involves categorisation skills, or it may require discriminating between two similar pieces of information. The different levels of proficiency associated with this process of comprehension can be measured by systematically varying the elements that contribute to the difficulty of the task.

Forming a broad general understanding

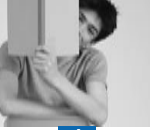
To *form a broad general understanding* of what has been read, a reader must consider the text as a whole or in a broad perspective. There are various assessment tasks in which readers are asked to form a broad general understanding. Students may demonstrate initial understanding by identifying the main topic or message or by identifying the general purpose or use of the text. Examples include tasks that require the reader to select or create a title or thesis for the text, to explain the order of simple instructions, or to identify the main dimensions of a graph or a table. Others include tasks that require the student to describe the main character, setting or milieu of a story, to identify a theme or message of a literary text, or to explain the purpose or use of a map or a figure.

Within this process some tasks might require the student to match a particular piece of text to the question. For example, this would happen when a theme or main idea is explicitly stated in the text. Other tasks may require the student to focus on more than one specific reference in the text – for instance, if the reader had to deduce the theme from the repetition of a particular category of information. Selecting the main idea implies establishing a hierarchy among ideas and choosing the most general and overarching. Such a task indicates whether the student can distinguish between key ideas and minor details, or can recognise the summary of the main theme in a sentence or title.

Developing an interpretation

Developing an interpretation requires readers to extend their initial impressions so that they develop a more specific or complete understanding of what they have read. Tasks in this category call for logical understanding; readers must process the organisation of information in the text. To do so, readers must demonstrate their understanding of cohesion even if they cannot explicitly state what cohesion is. In some instances, developing an interpretation may require the reader to process a sequence of just two sentences relying on local cohesion, which might even be facilitated by the presence of cohesive markers, such as the use of “first” and “second” to indicate a sequence. In more difficult instances (*e.g.* to indicate relations of cause and effect), there might not be any explicit markings.

Examples of tasks that might be used to assess this process include comparing and contrasting information, drawing inferences, and identifying and listing supporting evidence. “Compare and contrast” tasks require the student 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. 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.



Reflecting on and evaluating the content of a text

Reflecting on and evaluating the content of a text requires the reader to connect information in a text to knowledge from other 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 that with other sources of information, using both general and specific knowledge as well as the ability to reason abstractly.

Assessment tasks representative of this category of processing include providing evidence or arguments from outside the text, assessing the relevance of particular pieces of information or evidence, or drawing comparisons with moral or aesthetic rules (standards). The student might be asked to offer or identify alternative pieces of information that might strengthen an author's argument, or to evaluate the sufficiency of the evidence or information provided in the text.

The outside knowledge to which textual information is to be connected may come from the student's own knowledge, from other texts provided in the assessment, or from ideas explicitly provided in the question.

Reflecting on and evaluating the form of a text

Tasks in this category require readers to stand apart from the text, consider it objectively and evaluate its quality and appropriateness. Knowledge of such things as text structure, genre and register play an important role in these tasks. These features, which form the basis of an author's craft, figure strongly in understanding standards inherent in tasks of this nature. 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 nuances in language – for example, understanding when the choice of an adjective might colour interpretation.

Some examples of assessment tasks characteristic of *reflecting on and evaluating the form of a text* include determining the utility of a particular text for a specified purpose and evaluating an author's use of particular textual features in accomplishing a particular goal. The student may also be called upon to describe or comment on the author's use of style and to identify the author's purpose and attitude.

Distribution of tasks

Figure 2.3 shows the distribution of reading literacy tasks by each of the three subscales generated from the five reading processes (aspects) defined above. The largest category of tasks, which accounts for approximately 50% of the test, is represented by the two branches of Figure 2.2 that ask students to focus on relationships within a text. These tasks require students either to form a broad understanding or to develop an interpretation. They have been grouped together for reporting purposes into a single process called Interpreting texts. In PISA 2000, 2003 and 2006, the next largest category was made up of the 29% of the tasks that require students to demonstrate their skill at retrieving isolated pieces of information. Each of these processes – forming a broad understanding, retrieving information and developing an interpretation – focuses on the degree

Figure 2.3 • **Distribution of reading literacy tasks, by reading process (aspect)**

Reading as a major domain (PISA 2000)

Reading as a minor domain (PISA 2003 and 2006)

Reading process (aspect)	Percentage of tasks (PISA 2000)	Percentage of tasks (PISA 2003 and 2006)
Retrieving information	29	29
Interpreting texts	49	50
Reflection and evaluation	22	21
Total[1]	**100**	**100**

1. Data may not always add up to the totals indicated because of rounding.

to which the reader can understand and use information contained primarily within the text. The remaining tasks, approximately 20%, required students to reflect on either the content or information provided in the text or on the structure and form of the text itself.

Item types

The reading tasks in PISA are made up of various types, including multiple choice and various constructed tasks which require the students to write their answers rather than simply select from a number of given responses. The different types of tasks also require different marking. Figure 2.4 indicates that in PISA 2000, 2003 and 2006, around 43% of the reading literacy tasks in the PISA assessment were open-constructed response items which required judgement on the part of the marker. The remaining tasks consist of closed constructed-response items that require little judgement on the part of the marker, as well as simple multiple-choice items, for which students choose one of several alternative answers, and complex multiple-choice items, for which students choose more than one response.

This table also reveals that while multiple choice and open-constructed response items are represented across the processes, they are not distributed evenly.

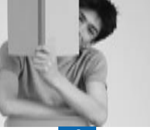
Figure 2.4 • **Distribution of reading literacy tasks, by reading process (aspect) and item type**

Reading as a major domain (PISA 2000)

Reading as a minor domain (PISA 2003 and 2006)

	Item types									
Process (aspect)	**Percentage of multiple-choice items**		**Percentage of complex multiple-choice items**		**Percentage of closed-constructed response items**		**Percentage of open-constructed response items**[1]		**Total**[2]	
Retrieving information	8	-	2	4	6	14	13	11	**29**	**29**
Interpreting texts	32	29	2	4	2	7	13	11	**49**	**50**
Reflection and evaluation	2	-	2	-	-	-	18	21	**22**	**21**
Total[2]	**42**	**29**	**6**	**7**	**9**	**21**	**44**	**43**	**100**	**100**

1. This category includes short-response items.
2. Data may not always add up to the total indicated because of rounding.



A larger percentage of multiple-choice items are associated with the two processes dealing with interpreting relationships within a text. This is shown in the second row of Figure 2.4. In contrast, while reflection and evaluation tasks account for around 20% in PISA 2000, 2003 and 2006, only 2% in 2000 are multiple choice. Of the reflection and evaluation tasks, around 20% are open-constructed response items that require judgement on the part of the marker.

Marking

Marking is relatively simple with dichotomously scored multiple-choice items: the student has either chosen the designated answer or not. Partial-credit models allow for more complex marking of items. Here, because some "wrong" answers are more close to being correct than others, students who provide an "almost right" answer receive partial credit. Psychometric models for such polytomous scoring are well-established and in some ways are preferable to dichotomous scoring, as they utilise more of the information in the responses. Interpretation of polytomous marking is more complex, however, as each task has several locations on the difficulty scale: one for the full-credit answer and others for each of the partial-credit answers. Partial-credit marking is used for some of the more complex constructed-response items in PISA.

SITUATIONS

The manner in which situation was defined was borrowed from the Council of Europe's work on language. Four situation variables were identified: *reading for private use, reading for public use, reading for work* and *reading for education*. While the intention of the PISA reading literacy assessment was to measure the kinds of reading that occur both within and outside classrooms, the manner in which situation was defined could not be based simply on where the reading activity is carried out. For example, textbooks are read both in schools and in homes, and the process and purpose of reading these texts differ little from one setting to another. Moreover, reading also involves the author's intended use, different types of content and the fact that others (*e.g.* teachers and employers) sometimes decide what should be read and for what purpose.

Thus, for the purpose of this assessment, situation can be understood as a general categorisation of texts based on the author's intended use, on the relationship with other persons implicitly or explicitly associated with the text, and on the general content. The sample texts were drawn from a variety of situations to maximise the diversity of content included in the reading literacy survey. Close attention was also paid to the origin of texts selected for inclusion in this survey. The goal was to reach a balance between the broad definition of *reading literacy* used in PISA and the linguistic and cultural diversity of participating countries. This diversity helped to ensure that no one group would be either advantaged or disadvantaged by the assessment content.

The four situation variables taken from the work of the Council of Europe are described as follows:

- *Reading for private use (personal):* This type of reading is carried out to satisfy an individual's own interests, both practical and intellectual. It also includes reading to maintain or develop personal connections to other people. Contents typically include personal letters, fiction, biography and informational texts read for curiosity, as a part of leisure or recreational activities.

- *Reading for public use:* This type of reading is carried out to participate in the activities of the wider society. It includes the use of official documents as well as information about public events. In general, these tasks are associated with more or less anonymous contact with others.
- *Reading for work (occupational):* While not all 15-year-olds will actually have to read at work, it is important to assess their readiness to move into the world of work since, in most countries, over 50% of them will be in the labour force within one to two years. The prototypical tasks of this type are often referred to as "reading to do" (Sticht, 1975; Stiggins, 1982) in that they are tied to the accomplishment of some immediate task.
- *Reading for education:* This type of reading is normally involved with acquiring information as part of a larger learning task. The materials are often not chosen by the reader, but assigned by a teacher. The content is usually designed specifically for the purpose of instruction. The prototypical tasks are those usually identified as "reading to learn" (Sticht, 1975; Stiggins, 1982).

Figure 2.5 shows the distribution of reading literacy tasks in the assessment across all four situations for two scenarios: when reading was a major domain (PISA 2000) and when it is a minor domain (PISA 2003 and 2006).

Figure 2.5 • **Distribution of reading literacy tasks, by situation**

Reading as a major domain (PISA 2000)
Reading as a minor domain (PISA 2003 and 2006)

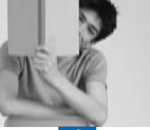
Situation	Percentage of tasks (PISA 2000)	Percentage of tasks (PISA 2003 and 2006)
Personal	20	21
Public	38	25
Occupational	14	25
Educational	28	29
Total	**100**	**100**

REPORTING OUTCOMES

Scaling the reading literacy tasks

The reading literacy tasks are constructed and administered to nationally representative samples of 15-year-old students in participating countries to ensure that the assessment provides the broadest possible coverage of *reading literacy* as defined here. However, no individual student can be expected to respond to the entire set of tasks. Accordingly, the survey is designed to give each student participating in the study a subset of the total pool of tasks, while at the same time ensuring that each of the tasks is administered to nationally representative samples of students. Summarising the performance of students across this entire pool of tasks thus poses a challenge.

One may imagine the reading literacy tasks arranged along a continuum in terms of difficulty for students and the level of skill required to answer each item correctly. The procedure used in PISA to capture this continuum of difficulty and ability is Item Response Theory (IRT). IRT is a mathematical model used for estimating the probability that a particular person will respond correctly to a given



task from a specified pool of tasks. This probability is modelled along a continuum which summarises both the proficiency of a person in terms of his or her ability and the complexity of an item in terms of its difficulty. This continuum of difficulty and proficiency is referred to as a "scale".

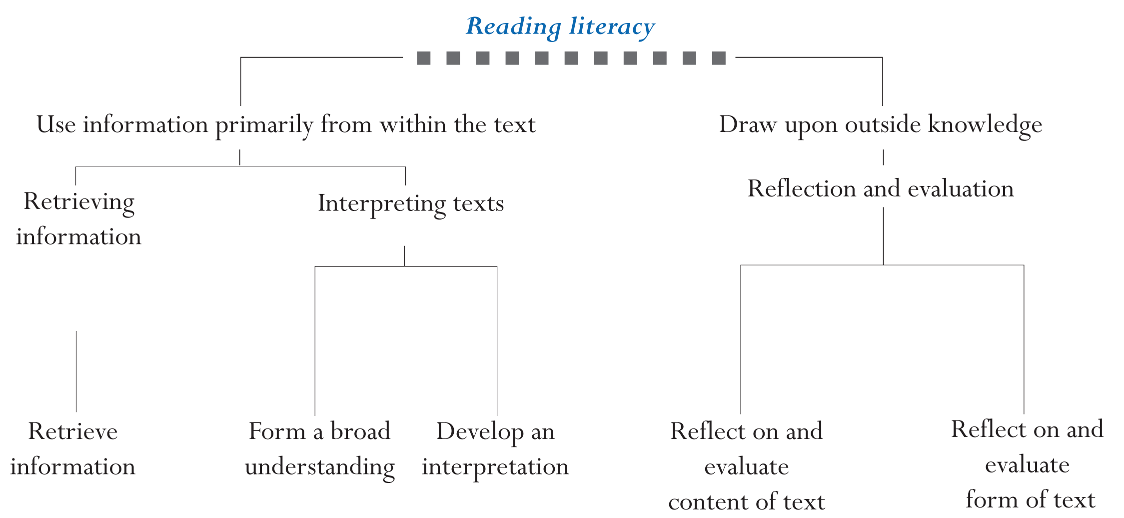
Reporting

PISA 2006 will follow the reporting scheme used in PISA 2000 and 2003, which reported outcomes in terms of a proficiency scale based on theory and interpretable in policy terms. The results of the reading literacy assessment were first summarised on a single composite reading literacy scale having a mean of 500 and a standard deviation of 100. In addition, student performance was also represented on five subscales: three process (aspect) subscales ("retrieving information", "interpreting texts", and "reflection and evaluation") (OECD, 2001) and two text format subscales (*continuous* and *non-continuous text*) (OECD, 2002). These five subscales make it possible to compare mean scores and distributions among subgroups and countries by various components of the reading literacy construct. Although there is a high correlation between these subscales, reporting results on each subscale may reveal interesting interactions among the participating countries. Where such features occur, they can be examined and linked to the curriculum and teaching methodology used. In some countries, the important question may be how to teach the current curriculum better. In others, the question may not only be how to teach but also what to teach.

The reading process (aspect) subscales

Figure 2.6 summarises the reading literacy tasks in terms of three processes. There are two reasons for reducing the number of processes from five to three for reporting purposes. The first is pragmatic. In 2003 and 2006, reading, as a minor domain, is restricted to about 30 items instead of the 141 that were used in 2000 when reading was a major domain. The amount of information, therefore, is insufficient to report trends over five process subscales. The second reason is conceptual. The three process subscales are based on the set of five processes shown in Figure 2.2. *Forming a broad understanding* and *developing an interpretation* have been grouped together in an "interpreting texts"

Figure 2.6 • **Relationship between the reading literacy framework and the process (aspect) subscales**



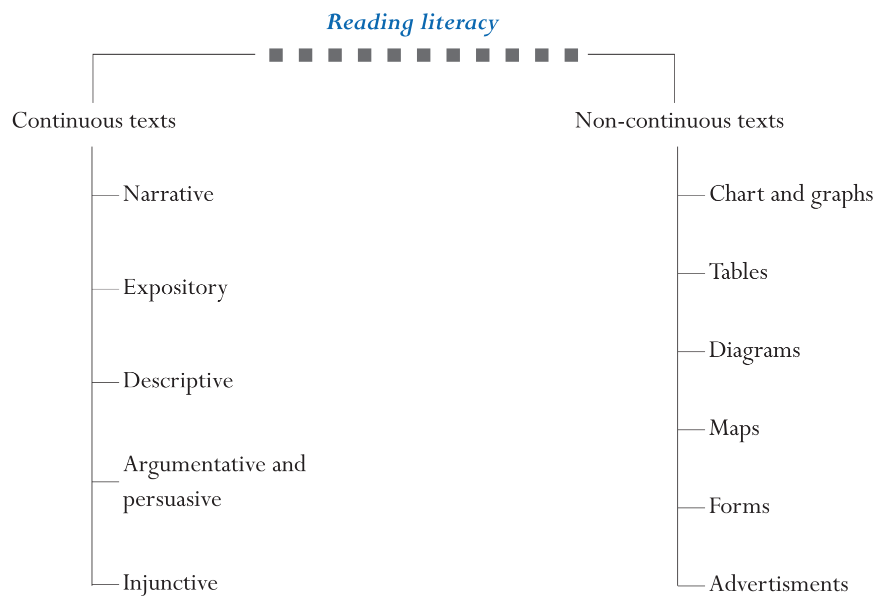
subscale because, in both, the reader processes information in the text: in the case of *forming a broad understanding*, the whole text and in the case of *developing an interpretation*, one part of the text in relation to another. *Reflecting on and evaluating the content of a text* and *reflecting on and evaluating the form of a text* have been collapsed into a single "reflection and evaluation" subscale because the distinction between reflecting on and evaluating content and reflecting on and evaluating form, in practice, was found to be somewhat arbitrary.

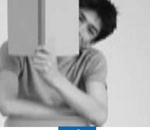
The text format subscales

PISA 2003 and 2006 also offer the possibility of providing results based on text format subscales, as reported in *Reading for Change: Performance and Engagement across Countries* (OECD, 2002). Figure 2.7 summarises the various text formats and the associated tasks along the two text format subscales. Organising the data in this way provides the opportunity to examine to what extent countries differ with respect to ability to deal with texts in different formats. In reporting results for 2000, two-thirds of the tasks were used to create the continuous text subscale while the remaining one-third of the tasks was used to create the non-continuous text subscale. There is a similar distribution of tasks between the two text formats in PISA 2003 and 2006.

The scores on the composite scale as well as on each of the five subscales represent varying degrees of proficiency. A low score indicates that a student has very limited knowledge and skills, while a high score indicates that a student has quite advanced knowledge and skills. Use of IRT makes it possible not only to summarise results for various subpopulations of students, but also to determine the relative difficulty of the reading literacy tasks included in the survey. In other words, just as individuals receive a specific value on a scale according to their performance in the assessment tasks, each task receives a specific value on a scale according to its difficulty, as determined by the performance of students across the various countries that participate in the assessment.

Figure 2.7 • **Relationship between the reading literacy framework and the text format subscales**





Building an item map

The complete set of reading literacy tasks used in PISA varies widely in text format, situation and task requirements, and hence also in difficulty. This range is captured through what is known as an item map. The item map provides a visual representation of the reading literacy skills demonstrated by students along the scales. The map should contain a brief description of a selected number of released assessment tasks along with their scale values. These descriptions take into consideration the specific skills the item is designed to assess and, in the case of open-ended tasks, the criteria used for judging the item correct. An examination of the descriptions provides some insight into the range of processes required of students and the proficiencies they need to demonstrate at various points along the reading literacy scales.

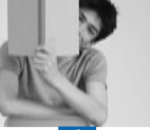
Figure 2.8 shows an example of an item map from PISA 2000. An explanation of how to interpret it may be useful. The score assigned to each item is based on the theory that someone at a given point on the scale is equally proficient in all tasks at that point on the scale. It was decided that, for the purposes of PISA, "proficiency" should mean that students at a particular point on the reading literacy scale would have a 62% chance of responding correctly to items at that point. For example, in Figure 2.8 an item appears at 421 on the composite scale. This means that students scoring 421 on the composite reading literacy scale will have a 62% chance of correctly answering items graded 421 on the scale. This does not mean that students receiving scores below 421 will always answer incorrectly. Rather, students scoring below 421 will be expected to answer correctly an item of that level of difficulty less than 62% of the time. Conversely, students having scores above 421 will have a greater than 62% chance of responding correctly. It should be noted that the item will also appear on a process subscale and on a format subscale as well as on the combined reading literacy scale. In this example, the item at 421 on the composite scale requires students to identify the purpose that two short texts have in common by comparing the main ideas in each of them. It is an interpretation item and thus appears on the interpreting texts scale as well as on the continuous texts scale.

Levels of reading literacy proficiency

Just as students within each country are sampled to represent the national population of 15-year-old students, each reading literacy task represents a class of tasks from the reading literacy domain. Hence, it represents proficiency in a type of processing and in dealing with a type of text that 15-year-old students should have acquired. One obvious question is, what distinguishes tasks at the lower end of the scale from those in the middle and upper ranges of the scale? Also, do tasks that fall around the same place on the scale share some characteristics that result in their having similar levels of difficulty? Even a cursory review of the item map reveals that tasks at the lower end of each scale differ from those at the higher end. A more careful analysis of the range of tasks along each scale provides indications of an ordered set of information-processing skills and strategies. Members of the reading expert group examined each task to identify a set of variables that seemed to influence its difficulty. They found that difficulty is in part determined by the length, structure and complexity of the text itself. However, they also noted that in most reading units (a unit being a text and a set of questions), the questions range across the reading literacy scale. This means that while the structure of a text contributes to the difficulty of an item, what the reader has to do with that text, as defined by the question or directive, interacts with the text and affects the overall difficulty.

Figure 2.8 • An example of a PISA 2000 item map

Composite item map	Types of process (Aspect)			Text format	
	Retrieving information	Interpreting	Reflection and evaluation	Continuous	Non-continuous
822 **HYPOTHESISE** about an unexpected phenomenon by taking account of outside knowledge along with all relevant information in a **COMPLEX TABLE** on a relatively unfamiliar topic. (Score 2)			○		○
727 **ANALYSE** several described cases and **MATCH** to categories given in a **TREE DIAGRAM**, where some of the relevant information is in footnotes. (Score 2)		○			○
705 **HYPOTHESISE** about an unexpected phenomenon by taking account of outside knowledge along with some relevant information in a **COMPLEX TABLE** on a relatively unfamiliar topic. (Score 1)			○		○
652 **EVALUATE** the ending of a LONG NARRATIVE in relation to its implicit theme or mood. (Score 2)			○	○	
645 **RELATE NUANCES OF LANGUAGE** in a **LONG NARRATIVE** to the main theme, in the presence of conflicting ideas. (Score 2)		○		○	
631 **LOCATE** information in a **TREE DIAGRAM** using information in a footnote. (Score 2)	○				○
603 **CONSTRUE** the meaning of a sentence by relating it to broad context in a **LONG NARRATIVE**.		○		○	
600 **HYPOTHESISE** about an authorial decision by relating evidence in a graph to the inferred main theme of **MULTIPLE GRAPHIC DISPLAYS**.			○		○
581 **COMPARE AND EVALUATE** the style of two open **LETTERS**.			○	○	
567 **EVALUATE** the ending of a **LONG NARRATIVE** in relation to the plot.			○	○	
542 **INFER AN ANALOGICAL RELATIONSHIP** between two phenomena discussed in an open **LETTER**.		○		○	
540 **IDENTIFY** the implied starting date of a **GRAPH**.	○				○
539 **CONSTRUE THE MEANING** of short quotations from a **LONG NARRATIVE** in relation to atmosphere or immediate situation. (Score 1)		○		○	
537 **CONNECT** evidence from **LONG NARRATIVE** to personal concepts in order to justify opposing points of view. (Score 2)			○	○	
529 **EXPLAIN** a character's motivation by linking events in a **LONG NARRATIVE**.		○		○	
508 **INFER THE RELATIONSHIP** between **TWO GRAPHIC DISPLAYS** with different conventions.		○			○
486 **EVALUATE** the suitability of a **TREE DIAGRAM** for particular purposes.			○		○
485 **LOCATE** numerical information in a **TREE DIAGRAM**.	○				○
480 **CONNECT** evidence from **LONG NARRATIVE** to personal concepts in order to justify a single point of view. (Score 1)			○	○	
478 **LOCATE AND COMBINE** information in a **LINE GRAPH** and its introduction to infer a missing value.	○				○
477 **UNDERSTAND** the structure of a **TREE DIAGRAM**.		○			○
473 **MATCH** categories given in a **TREE DIAGRAM** to described cases, when some of the relevant information is in footnotes.		○			○
447 **INTERPRET** information in a single paragraph to understand the setting of a **NARRATIVE**.		○		○	
445 Distinguish between variables and **STRUCTURAL FEATURES** of a **TREE DIAGRAM**.			○		○
421 **IDENTIFY** the common **PURPOSE** of **TWO SHORT TEXTS**.		○		○	
405 **LOCATE** pieces of explicit information in a **TEXT** containing strong organizers.	○			○	
397 Infer the **MAIN IDEA** of a simple **BAR GRAPH** from its title.		○			○
392 **LOCATE** a literal piece of information in a **TEXT** with clear text structure.	○			○	
367 **LOCATE** explicit information in a short, specified section of a **NARRATIVE**.	○			○	
363 **LOCATE** an explicitly stated piece of information in a **TEXT** with headings.	○			○	
356 **RECOGNISE THEME** of an article having a clear subheading and considerable redundancy.		○		○	



The members of the reading expert group and test developers identified a number of variables that can influence the difficulty of any reading literacy task. One salient factor is the process involved in retrieving information, developing an interpretation or reflecting on what has been read. Processes range in complexity and sophistication from making simple connections between pieces of information, to categorising ideas according to given criteria, and to critically evaluating and hypothesising about a section of text. In addition to the process called for, the difficulty of retrieving information tasks varies with the number of pieces of information to be included in the response, the number of criteria which the information must satisfy, and whether or not what is retrieved needs to be sequenced in a particular way. In the case of interpretative and reflective tasks, the amount of a text that needs to be assimilated is an important factor affecting difficulty. In items that require reflection on the reader's part, difficulty is also conditioned by the familiarity or specificity of the knowledge that must be drawn on from outside the text. In all processes of reading, the difficulty of the task depends on how prominent the required information is, how much competing information is present, and whether or not the reader is explicitly directed to the ideas or information required to complete the task.

In an attempt to capture this progression of complexity and difficulty in PISA 2000, the composite reading literacy scale and each of the subscales were divided into five levels:

Level	Score points on the PISA scale
1	335 to 407
2	408 to 480
3	481 to 552
4	553 to 625
5	More than 625

Expert panels judged that the tasks within each level of *reading literacy* shared many of the same task features and requirements, and differed in systematic ways from tasks at higher or lower levels. As a result, these levels appear to be a useful way to explore the progression of reading literacy demands within each scale. This progression is summarised in Figure 2.9. This process was undertaken for *mathematical literacy* in PISA 2003 and will be undertaken for *scientific literacy* in 2006.

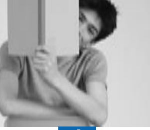
Interpreting the reading literacy levels

Not only does each level represent a range of tasks and associated knowledge and skills, it also represents a range of proficiencies demonstrated by students. As mentioned previously, the reading literacy levels were initially set by the members of the reading expert group to represent a set of tasks with shared characteristics. These levels also have shared statistical properties. The average student within each level can be expected to successfully perform the average task within that level 62% of the time. In addition, the width of each level is in part determined by the expectation that a student at the lower end of any level will score 50% on any hypothetical test made up of items randomly selected from that level.

Since each reading literacy scale represents a progression of knowledge and skills, students at a particular level not only demonstrate the knowledge and skills associated with that particular level but the proficiencies associated with the lower levels as well. Thus the knowledge and skills assumed

Figure 2.9 • Reading literacy levels map

<table>
<tr><th></th><th>Retrieving information</th><th>Interpreting texts</th><th>Reflection and evaluation</th></tr>
<tr><td>Level 5</td><td>Locate and possibly sequence or combine multiple pieces of deeply embedded information, some of which may be outside the main body of the text. Infer which information in the text is relevant to the task. Deal with highly plausible and/or extensive competing information.</td><td>Either construe the meaning of nuanced language or demonstrate a full and detailed understanding of a text.</td><td>Critically evaluate or hypothesise, drawing on specialised knowledge. Deal with concepts that are contrary to expectations and draw on a deep understanding of long or complex texts.</td></tr>
<tr><td></td><td colspan="3">Continuous texts: Negotiate texts whose discourse structure is not obvious or clearly marked, in order to discern the relationship of specific parts of the text to its implicit theme or intention.
Non-continuous texts: Identify patterns among many pieces of information presented in a display which may be long and detailed, sometimes by referring to information external to the display. The reader may need to realise independently that a full understanding of the section of text requires reference to a separate part of the same document, such as a footnote.</td></tr>
<tr><td>Level 4</td><td>Locate and possibly sequence or combine multiple pieces of embedded information, each of which may need to meet multiple criteria, in a text with unfamiliar context or form. Infer which information in the text is relevant to the task.</td><td>Use a high level of text-based inference to understand and apply categories in an unfamiliar context, and to construe the meaning of a section of text by taking into account the text as a whole. Deal with ambiguities, ideas that are contrary to expectation and ideas that are negatively worded.</td><td>Use formal or public knowledge to hypothesise about or critically evaluate a text. Show accurate understanding of long or complex texts.</td></tr>
<tr><td></td><td colspan="3">Continuous texts: Follow linguistic or thematic links over several paragraphs, often in the absence of clear discourse markers, in order to locate, interpret or evaluate embedded information or to infer psychological or metaphysical meaning.
Non-continuous texts: Scan a long, detailed text in order to find relevant information, often with little or no assistance from organisers such as labels or special formatting, and to locate several pieces of information to be compared or combined.</td></tr>
<tr><td>Level 3</td><td>Locate, and in some cases recognise, the relationship between pieces of information, each of which may need to meet multiple criteria. Deal with prominent competing information.</td><td>Integrate several parts of a text in order to identify a main idea, understand a relationship or construe the meaning of a word or phrase. Compare, contrast or categorise taking many criteria into account. Deal with competing information.</td><td>Make connections or comparisons, give explanations, or evaluate a feature of text. Demonstrate a detailed understanding of the text in relation to familiar, everyday knowledge, or draw on less common knowledge.</td></tr>
<tr><td></td><td colspan="3">Continuous texts: Use conventions of text organisation, where present, and follow implicit or explicit logical links such as cause and effect relationships across sentences or paragraphs in order to locate, interpret or evaluate information.
Non-continuous texts: Consider one display in the light of a second, separate document or display, possibly in a different format, or combine several pieces of spatial, verbal and numeric information in a graph or map to draw conclusions about the information represented.</td></tr>
<tr><td>Level 2</td><td>Locate one or more pieces of information, each of which may be required to meet multiple criteria. Deal with competing information.</td><td>Identify the main idea in a text, understand relationships, form or apply simple categories, or construe meaning within a limited part of the text when the information is not prominent and low-level inferences are required.</td><td>Make a comparison or connections between the text and outside knowledge, or explain a feature of the text by drawing on personal experience and attitudes.</td></tr>
<tr><td></td><td colspan="3">Continuous texts: Follow logical and linguistic connections within a paragraph in order to locate or interpret information; or synthesise information across texts or parts of a text in order to infer the author's purpose.
Non-continuous texts: Demonstrate a grasp of the underlying structure of a visual display such as a simple tree diagram or table, or combine two pieces of information from a graph or table.</td></tr>
<tr><td>Level 1</td><td>Locate one or more independent pieces of explicitly stated information, typically meeting a single criterion, with little or no competing information in the text.</td><td>Recognise the main theme or author's purpose in a text about a familiar topic, when the required information in the text is prominent.</td><td>Make a simple connection between information in the text and common, everyday knowledge.</td></tr>
<tr><td></td><td colspan="3">Continuous texts: Use redundancy, paragraph headings or common print conventions to form an impression of the main idea of the text, or to locate information stated explicitly within a short section of text.
Non-continuous texts: Focus on discrete pieces of information, usually within a single display such as a simple map, a line graph or a bar graph that presents only a small amount of information in a straightforward way, and in which most of the verbal text is limited to a small number of words or phrases.</td></tr>
</table>



at each level build on and encompass the proficiencies laid down in the next lower level. This means that a student who is judged to be at Level 3 on a reading literacy scale is proficient not only in Level 3 tasks but also in Level 1 and 2 tasks. This also means that students who are at Levels 1 and 2 will be expected to get the average Level 3 item correct less than 50% of the time. Put another way, they will be expected to score less than 50% on a test made up of items drawn from Level 3.

Figure 2.10 shows the probability that individuals performing at selected points along the combined reading literacy scale will give a correct response to tasks of varying difficulty. One is a Level 1 task, one is a Level 3 task, and the third task receives two score points: one at Level 4 and the other at Level 5. It is readily seen here that a student with a score of 298, who is estimated to be below Level 1, has only a 43% chance of responding correctly to the Level 1 task that is at 367 on the reading literacy scale. This person has only a 14% chance of responding to the item from Level 3 and almost no chance of responding correctly to the item from Level 5. Someone with a proficiency of 371, in the middle of Level 1, has a 63% chance of responding to the item at 367, but only slightly more than one chance in four of responding correctly to the task at 508, and only a seven% chance of responding correctly to the task selected from Level 5. In contrast, someone at Level 3 would be expected to respond correctly 89% of the time to tasks at 367 on the reading literacy scale, and 64% of the time to tasks at 508, near the middle of Level 3. However, he or she would only have just over one chance in four (27%) of correctly responding to items from the middle of Level 5. Finally, a student at Level 5 is expected to respond correctly most of the time to almost all the tasks. As shown in Figure 2.10, a student having a score of 662 on the combined reading literacy scale has a 98% chance of answering the task at 367 correctly, a 90% chance of answering the item at Level 3 (508) correctly and a 65% of responding correctly to the task selected from near the centre of Level 5 (652).

Figure 2.10 also implicitly raises questions concerning the highest and lowest designated levels. Even though the top of the reading literacy scale is unbounded, it can be stated with some certainty that students of extremely high proficiency are capable of performing tasks characterised by the highest level of proficiency. There is more of an issue for students who are at the bottom end of the reading literacy scale. Level 1 begins at 335, yet a certain percentage of students in each country is estimated to be below this point on the scale. While there are no reading literacy tasks with a scale

Figure 2.10 • **Probability of responding correctly to selected tasks of varying difficulty for students with varying levels of proficiency**

Students with varying levels of proficiency	Selected tasks of varying difficulty: Level 1 item at 367 points	Level 3 item at 508 points	Level 4 item at 567 points	Level 5 item at 652 points
Below Level 1 (Proficiency of 298 points)	43	14	8	3
Level 1 (Proficiency of 371 points)	63	27	16	7
Level 2 (Proficiency of 444 points)	79	45	30	14
Level 3 (Proficiency of 517 points)	89	64	48	27
Level 4 (Proficiency of 589 points)	95	80	68	45
Level 5 (Proficiency of 662 points)	98	90	82	65

