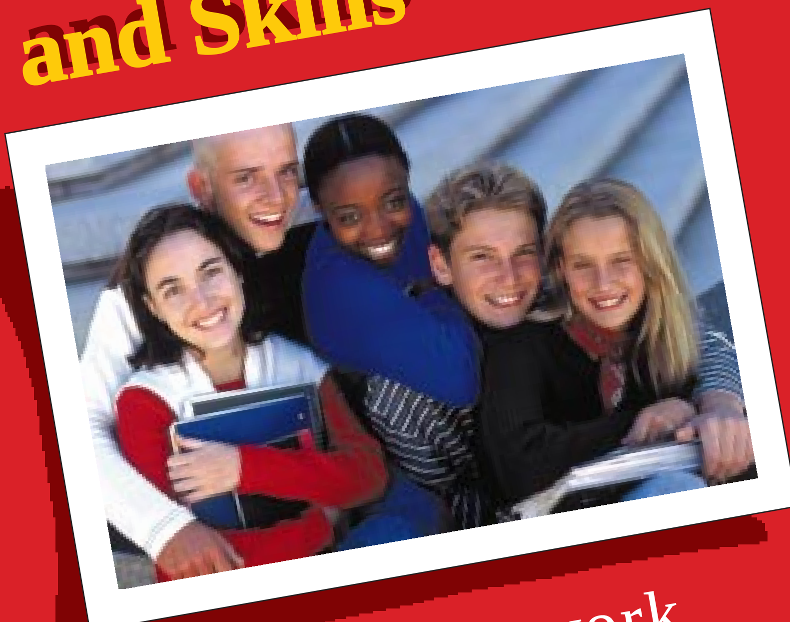
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Measuring Student Knowledge and Skills



A New Framework for Assessment

OECD PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT

MEASURING STUDENT KNOWLEDGE AND SKILLS

A New Framework for Assessment

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Pursuant to Article 1 of the Convention signed in Paris on 14th December 1960, and which came into force on 30th September 1961, the Organisation for Economic Co-operation and Development (OECD) shall promote policies designed:

– to achieve the highest sustainable economic growth and employment and a rising standard of living in Member countries, while maintaining financial stability, and thus to contribute to the development of the world economy;

– to contribute to sound economic expansion in Member as well as non-member countries in the process of economic development; and

– to contribute to the expansion of world trade on a multilateral, non-discriminatory basis in accordance with international obligations.

The original Member countries of the OECD are Austria, Belgium, Canada, Denmark, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The following countries became Members subsequently through accession at the dates indicated hereafter: Japan (28th April 1964), Finland (28th January 1969), Australia (7th June 1971), New Zealand (29th May 1973), Mexico (18th May 1994), the Czech Republic (21st December 1995), Hungary (7th May 1996), Poland (22nd November 1996) and Korea (12th December 1996). The Commission of the European Communities takes part in the work of the OECD (Article 13 of the OECD Convention).

Publié en français sous le titre :
MESURER LES CONNAISSANCES ET COMPÉTENCES DES ÉLÈVES
Un nouveau cadre d'évaluation

Photo Credit: PIX/Denis Boissavy

FOREWORD

The OECD's International Programme for Student Assessment (PISA) represents a new commitment by the governments of OECD 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.

The design and implementation of the surveys is undertaken under the direction of the OECD Secretariat, through an international consortium led by the Australian Council for Educational Research (ACER). Other partners in this consortium include the Netherlands National Institute for Educational Measurement, the Service de pédagogie expérimentale de l'Université de Liège, and WESTAT.

This publication sets out the conceptual framework underlying the OECD/PISA assessments: it defines each domain to be assessed and explains what will be assessed and how. It also describes the context and constraints within which the OECD/PISA assessments are placed. It is published on the responsibility of the Secretary-General of the OECD.

ACKNOWLEDGEMENTS

The PISA assessment frameworks were developed by expert panels under the direction of Raymond Adams from the ACER. The reading expert panel was chaired by Dr. Irwin Kirsch of Educational Testing Service, the mathematics panel was chaired by Professor Jan de Lange of the University of Utrecht, and the science expert panel was chaired by Professor Wynne Harlen of the Scottish Council for Research in Education. The members of the functional expert groups are listed in Appendix 1. The frameworks have also been informed by reviews from expert panels in each of the participating countries. They were adopted in December 1998 by OECD governments, through the Board of Participating Countries. The publication was prepared by the Statistics and Indicators Division of the Directorate for Education, Employment, Labour and Social affairs, principally Andreas Schleicher.

TABLE OF CONTENTS

INTRODUCTION

How well are young adults prepared to meet the challenges of the future? Are they able to analyse, reason and communicate their ideas effectively? Do they have the capacity to continue learning throughout life? Parents, students, the public and those who run education systems need to know.

Many education systems monitor student learning to provide some answers to these questions. Comparative international analyses can extend and enrich the national picture by establishing the levels of performance being achieved by students in other countries and by providing a larger context within which to interpret national results. They can provide direction for schools' instructional efforts and for students' learning as well as insights into curriculum strengths and weaknesses. Coupled with appropriate incentives, they can motivate students to learn better, teachers to teach better and schools to be more effective. They also provide tools to allow central authorities to monitor achievement levels even when administration is devolved and schools are being run in partnership with communities.

Governments and the general public need solid and internationally comparable evidence of educational outcomes. In response to this demand, the Organisation for Economic Co-operation and Development (OECD) has launched the Programme for International Student Assessment (PISA). OECD/PISA will produce policy-oriented and internationally comparable indicators of student achievement on a regular and timely basis. The assessments will focus on 15-year-olds, and the indicators are designed to contribute to an understanding of the extent to which education systems in participating countries are preparing their students to become lifelong learners and to play constructive roles as citizens in society.

OECD/PISA represents a new commitment by the governments of OECD countries to monitor the outcomes of education systems in terms of student achievement, within a common framework that is internationally agreed. While it is expected that many individuals in participating countries, including professionals and lay-persons, will use the survey results for a variety of purposes, the primary reason for developing and conducting this large-scale international assessment is to provide empirically grounded information which will inform policy decisions.

The results of the OECD assessments, to be published every three years along with other indicators of education systems, will allow national policy makers to compare the performance of their education systems with those of other countries. They will also help to focus and motivate educational reform and school improvement, especially where schools or education systems with similar inputs achieve markedly different results. Further, they will provide a basis for better assessment and monitoring of the effectiveness of education systems at the national level.

OECD/PISA is a collaborative process. It brings together scientific expertise from the participating countries and is steered jointly by their governments on the basis of shared, policy-driven interests. Following an overview of the design of OECD/PISA and a summary of the most important features of the assessments, there is a description of the nature of this collaboration, and how it is being used to develop the frameworks that will define the OECD assessments.

The remaining sections of this publication then set out the conceptual framework underlying the OECD/PISA assessments: they define each domain to be assessed and explain *what* will be assessed and *how*. They also describe the context and constraints within which the OECD/PISA assessments are placed.

What is OECD/PISA? A summary of key features

Basics

- An internationally standardised assessment, jointly developed by participating countries and administered to 15-year-olds in groups in their schools.
- Administered in 32 countries, of which 28 are members of the OECD.
- Between 4 500 and 10 000 students will typically be tested in each country.

Content

- PISA covers three domains: reading literacy, mathematical literacy and scientific literacy.
- PISA aims to define each domain not merely in terms of mastery of the school curriculum, but in terms of important knowledge and skills needed in adult life. The assessment of cross-curriculum competencies is an integral part of PISA.
- Emphasis is placed on the mastery of processes, the understanding of concepts and the ability to function in various situations within each domain.

Methods

- Pencil and paper tests are used, with assessments lasting a total of 2 hours for each student.
- Test items are a mixture of multiple-choice test items and questions requiring the student 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 7 hours of test items is included, with different students taking different combinations of the test items.
- Students answer a background questionnaire which takes 20-30 minutes to complete, providing information about themselves. School principals are given a 30-minute questionnaire asking about their schools.

Assessment cycle

- The first assessment will take place in 2000, with first results published in 2001, and will continue thereafter, in three-year cycles.
- Each cycle looks in depth at a “major” domain, to which two-thirds of testing time are devoted; the other two 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 students at the end of compulsory schooling.
- Contextual indicators relating results to student and school characteristics.
- Trend indicators showing how results change over time.

THE DESIGN OF OECD/PISA 2000

Basic features of OECD/PISA

The OECD assessments in 2000 will cover reading literacy, mathematical literacy and scientific literacy. Students will also be responding to a background questionnaire and supporting information will be gathered from the school authorities. The first OECD assessments will be administered in 2000 and results will become available from 2001 onwards. Thirty-two countries, including 28 OECD Member countries, plan to take part in OECD/PISA. Together, these countries represent more than a quarter of the world population, more than that covered in any international educational assessment to date.

Since the aim of OECD/PISA is to assess the cumulative yield of education systems at an age where schooling is still largely universal, testing will focus on 15-year-olds enrolled in both school-based and work-based educational programmes. Between 4 500 and 10 000 students 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.

Although the domains of reading literacy, mathematical literacy and scientific literacy correspond to school subjects, the OECD assessments will not primarily examine how well students have mastered the specific curriculum content. Rather, they aim at assessing the extent to which young people have acquired the wider knowledge and skills in these domains that they will need in adult life. The assessment of cross-curricular competencies has, therefore, been made an integral part of OECD/PISA. The most important reasons for this broadly oriented approach to assessment are as follows:

– First, although specific knowledge acquisition is important in school learning, the application of that knowledge in adult life depends crucially on the individual's acquisition of broader concepts and skills. 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 important than the ability to answer familiar textbook questions when it comes to deploying mathematical skills in everyday life. In science, having specific knowledge, such as the names of specific plants and animals, is of less value than an understanding of broad concepts and topics such as energy consumption, biodiversity and human health in thinking about the issues of science under debate in the adult community.

– Second, a focus on curriculum content would, in an international setting, restrict attention to curriculum elements common to all, or most, countries. This would force many compromises and result in an assessment that was too narrow to be of value for governments wishing to learn about the strengths and innovations in the education systems of other countries.

– Third, there are broad, general skills that it is essential for students to develop. These 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 cross-curricular focus.

Underlying OECD/PISA is a dynamic model of lifelong learning in which new knowledge and skills necessary for successful adaptation to changing circumstances are continuously acquired over the life cycle. Students cannot learn in school everything they will need to know in adult life. What they must acquire is the prerequisites for successful learning in future life. These prerequisites are of both a cognitive and a motivational nature. Students must become able to organise and regulate their own learning, to learn independently and in groups, and to overcome difficulties in the learning process. This requires them to be aware of their own thinking processes and learning strategies and methods. Moreover, further learning and the acquisition of additional knowledge will increasingly occur in situations in which people work together and are dependent on one another. To assess these aspects, the development of an instrument that seeks information on self-regulated learning is being explored as part of the OECD/PISA 2000 assessment.

OECD/PISA is not a single cross-national assessment of the reading, mathematics and science skills of 15-year-olds. It is an on-going programme of assessment that will gather data from each of these domains every three years. Over the longer term, this will lead to the development of a body of information for monitoring trends in the knowledge and skills of students in the various countries as well as in different demographic sub-groups of each country. On each occasion, one domain will be tested in detail, taking up nearly two-thirds of the total testing time. This “major” domain will be reading literacy in 2000, mathematical literacy in 2003 and scientific literacy in 2006. This cycle will provide a thorough analysis of achievement in each area every nine years, and a “check-up” every three.

The total time spent on the tests by each student will be two hours but information will be obtained on almost seven hours’ worth of test items. The total set of questions will be packaged into a number of different groups. Each group will be 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 will also spend 20 minutes answering questions for the context questionnaire.

The assessments will provide various types of indicators:

– basic indicators providing a baseline profile of the knowledge and skills of students;
– contextual indicators, showing how such skills relate to important demographic, social, economic and educational variables;
– indicators on trends that will emerge from the on-going, cyclical nature of the data collection and that will show changes in outcome levels, changes in outcome distributions and changes in relationships between student-level and school-level background variables and outcomes over time.

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

The countries participating in the first OECD/PISA survey cycle are: Australia, Austria, Belgium, Brazil, Canada, China, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Latvia, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Russian Federation, Spain, Sweden, Switzerland, the United Kingdom and the United States.

How OECD/PISA is different from other international assessments

OECD/PISA is not the first international comparative survey of student achievement. Others have been conducted over the past 40 years, primarily by the International Association for the Evaluation of Educational Achievement (IEA) and by the Education Testing Service’s International Assessment of Educational Progress (IAEP). The quality and scope of these surveys have greatly improved over the years but they provide only partial and sporadic information about student achievement in limited subject areas. The three science and mathematics surveys conducted by the IEA provide some indication of how things have changed over 30 years but the view is limited by the restricted numbers of countries participating in the early surveys and by limitations on the extent to which the tests can be compared.

More importantly, though, 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, regardless of how significant that part of the curriculum is for the countries involved.

OECD/PISA is taking a different approach in a number of important respects which make it distinctive:

– Its *origin*: it is governments that have taken the initiative and whose policy interests the survey will be designed to serve.

– Its *regularity*: the commitment to cover multiple assessment domains, with updates every three years, will make it possible for countries to regularly and predictably monitor 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 which 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 and ambitious novel feature of OECD/PISA. It would be arbitrary to draw too precise a distinction between "school" skills and "life" skills, since schools have always aimed to prepare young people for life, but the distinction is important. 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 one's ideas and understanding to situations encountered in life. OECD/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, OECD/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 of 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 the stock of human capital or human skill base 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, OECD/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. Such authentic outcome measures are needed if schools and education systems are to be encouraged to focus on modern challenges.

What is being assessed in each domain

Table 1 summarises the structure of each of the three OECD/PISA domains, giving the definition of each domain and the dimensions that characterise the test items.

The definitions of all three domains place an emphasis on 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

Table 1. **Summary of PISA dimensions**

Domain	Reading literacy	Mathematical literacy	Scientific literacy
Definition	***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.***	***Identifying, understanding and engaging in mathematics and making well-founded judgements about the role that mathematics plays, as needed for an individual's current and future life as a constructive, concerned and reflective citizen.***	***Combining scientific knowledge with the drawing of evidence-based conclusions and developing hypotheses in order to understand and help make decisions about the natural world and the changes made to it through human activity.***
Components/ dimensions of the domain	Reading different kinds of *text:* continuous prose sub-classified by type (*e.g.* description, narration) and documents, sub-classified by structure.	Mathematical *content* – primarily mathematical "big ideas". In the first cycle these are change and growth, and space and shape. In future cycles chance, quantitative reasoning, uncertainty and dependency relationships will also be used.	*Scientific concepts – e.g.* energy conservation, adaptation, decomposition – chosen from the major fields of physics, biology, chemistry etc. where they are applied in matters to do with the use of energy, the maintenance of species or the use of materials.
	Performing different kinds of reading *tasks*, such as retrieving specific information, developing an interpretation or reflecting on the content or form of the text.	Mathematical *competencies*, *e.g.* modelling, problem-solving; divided into three classes: *i)* carrying out procedures, *ii)* making connections and *iii)* mathematical thinking and generalisation.	*Process skills – e.g.* identifying evidence, drawing, evaluating and communicating conclusions. These do not depend on a pre-set body of scientific knowledge, but cannot be applied in the absence of scientific content.
	Reading texts written for different *situations, e.g.* for personal interest, or to meet work requirements.	Using mathematics in different situations, *e.g.* problems that affect individuals, communities or the whole world.	Using science in different situations, *e.g.* problems that affect individuals, communities or the whole world.

decision-making processes. In the more complex tasks in OECD/PISA, students will be asked to reflect on material, not just to answer questions that have single "correct" answers.

In order to operationalise these definitions, each domain is described in terms of three dimensions. These correspond roughly to:

– the *content* or *structure* of knowledge that students need to acquire in each domain;

– a range of *processes* that need to be performed, which require various cognitive skills; and

– the *situation* or *context* in which knowledge and skills are applied or drawn on.

The idea is to assess students across a range of skills required for a variety of tasks that anyone might have to perform. However, it should be emphasised that the relative importance of the three dimensions still needs to be explored. Field trials in 1999 will assess a large number of questions on the three dimensions, with varying characteristics, before a decision is reached about the most useful range of characteristics and the manner in which to translate performance in various items into aggregate scores.

Within the common framework of the three dimensions, each domain defines its various dimensions in particular ways. There is an important difference between reading literacy on the one hand, and science and mathematics on the other. The former is itself a skill that cuts across the curriculum, particularly in secondary education, and does not have any obvious "content" of its own. Although some formal understanding of structural features such as sentence structure may be important, such knowledge cannot be compared, for example, to the mastery of a range of scientific principles or concepts.

The following are the main aspects of each of the three assessment domains summarised in Table 1.

Reading literacy is defined in terms of individuals' ability to use 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 OECD/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:

- First, the *form of reading material*, or text. Many student reading assessments have focused on "continuous texts" or prose organised in sentences and paragraphs. OECD/PISA will in addition introduce "non-continuous texts" which 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 adult life, and that it is not sufficient to be able to read a limited number of text types typically encountered in school.
- Second, the *type of reading task*. This corresponds at one level to the various cognitive skills that are needed to be an effective reader, and at another, to the characteristics of questions set in the assessment. Students will not be assessed on the most basic reading skills, as it is assumed that most 15-year-olds will have acquired these. Rather, they will be 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.
- Third, the *use for which the text was constructed* – its context or situation. 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. An important reason for making these distinctions is that some groups may perform better in one reading situation than in another, in which case it is desirable to include a range of types of reading in the assessment items.

Mathematical literacy is defined in terms of the individual's understanding of the role of mathematics and the capacity to engage in this discipline in ways that meet his or her needs. This puts the emphasis on the capacity to pose and solve mathematical problems rather than to perform specified mathematical operations. Mathematical literacy is assessed in relation to:

- First, the *content* of mathematics, as defined mainly in terms of mathematical "big ideas" (such as chance, change and growth, space and shape, quantitative reasoning, uncertainty and dependency relationships) and only secondarily in relation to "curricular strands" (such as numbers, algebra and geometry). A representative, rather than a comprehensive range of the main concepts underlying mathematical thinking have been chosen for OECD/PISA; these have been narrowed down further for the first cycle of the assessment – in which mathematics is a "minor" domain – to two big ideas: change and growth and space and shape. These allow a wide representation of aspects of the curriculum without undue focus on number skills.
- Second, the *process* of mathematics as defined by general mathematical competencies. These include the use of mathematical language, modelling and problem-solving skills. The idea is not, however, to separate out such skills 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 three "competency classes" defining the type of thinking skill needed. The first class consists of simple computations or definitions of the type most familiar in conventional mathematics assessments. The second requires connections to be made to solve straightforward problems. The third competency class consists of mathematical thinking, generalisation and insight, and requires students to engage in analysis, to identify the mathematical elements in a situation and to pose their own problems.
- Third, the *situations* in which mathematics is used. The framework identifies five situations: personal, educational, occupational, public and scientific. In the case of mathematics this dimension is considered to be less important than process or content.

Scientific literacy is defined in terms of being able to use scientific knowledge and processes not just to understand the natural world but to participate in decisions that affect it. Scientific literacy is assessed in relation to:

- First, *scientific concepts*, which constitute the links aiding understanding of related phenomena. In OECD/PISA, whilst the concepts are the familiar ones relating to physics, chemistry, biological sciences and earth and space sciences, they will need to be applied to the content of the items and not just recalled. The main item content will be selected from within three broad areas of application: science in life and health, science in earth and environment and science in technology.

– Second, *scientific processes*, centred on the ability to acquire, interpret and act upon evidence. Five such processes present in OECD/PISA relate to: *i)* the recognition of scientific questions; *ii)* the identification of evidence; *iii)* the drawing of conclusions; *iv)* the communication of these conclusions; and *v)* the demonstration of understanding of scientific concepts. Only the last of these requires a pre-set body of scientific knowledge, yet since no scientific process can be "content free", the OECD/PISA science questions will always require an understanding of concepts in the areas mentioned above.

– Third, *scientific situations*, selected mainly from people's everyday lives rather than from the practice of science in a school classroom or laboratory, or the work of professional scientists. The science situations in OECD/PISA are defined as matters relating to the self and the family, to the wider community, to the whole world and to the historical evolution of scientific understanding.

How the assessment will take place and how results will be reported

For reasons of feasibility, OECD/PISA 2000 will consist of pencil and paper instruments. Other forms of assessments will be actively explored for subsequent cycles.

The assessment will be made up of items of a variety of types. Some items will be "closed" – that is, they will require students to select or produce simple responses that can be directly compared with a single correct answer. Others will be more open, requiring students to provide more elaborate responses designed to measure broader constructs than those captured by other, more traditional surveys. The assessment of higher-order skills, often through open-ended problems, will be an important innovation in OECD/PISA. The extent to which this type of exercise will be used will depend on how robust the methodology proves to be in the field trial and how consistent a form of marking can be developed. The use of open-ended items is likely to grow in importance in successive OECD/PISA cycles, from a relatively modest start in the first survey cycle.

In most cases the assessments will consist of sets of items relating to a common text, stimulus or theme. This is an important feature that allows questions to go into greater depth than would be the case 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.

Overall, the items in OECD/PISA will look quite different from those used, for example, in earlier international assessments such as the Third International Mathematics and Science Study (IEA/TIMSS), which concentrated on short multiple-choice questions based on what had been learned at school. For example, some science questions may only require straightforward knowledge (*e.g.* asking students to specify how many legs and body parts insects have) or simple manipulation of knowledge (*e.g.* students have to work out whether a metal, wooden or plastic spoon would feel hottest after being placed in hot water for 15 seconds). OECD/PISA items, on the other hand, generally require the combination of a variety of knowledge and competencies, and sometimes (such as in "Example Item 4" in the scientific literacy framework – see Figure 18) an active evaluation of decisions for which there is no single correct answer.

OECD/PISA results will be reported in terms of the level of performance on scales of achievement in each domain. The calibration of the tasks in the tests on to scales will provide a language with which to describe the competencies exhibited by students performing at different levels. That is, it will be possible to say what students at any level on each scale know and are able to do that those below them cannot. The inclusion of items that require higher order thinking skills as well as others that entail relatively simple levels of understanding will ensure that the scales cover a wide range of competencies.

An important issue will be whether the levels of literacy in each domain should be reported on more than one scale. Can a person's competencies be easily aggregated and placed on a specific level, or is it more useful to describe them as having reached different levels in relation to different aspects? That will depend on two things, which will become more evident in the field trials: first, the extent to which an individual's performance in one kind of question correlates with performance in another, and the patterns of any differences in performance across particular dimensions; and second, the feasibility, given the number of items that can be included in the assessment, of reporting on more than one scale in each domain. Each scale corresponds to one type of score that will be assigned to students; having more than one scale therefore implies that students will be assigned multiple scores reflecting different aspects of the domain. The most likely scenario is that in the major domain that contains the most questions (reading in OECD/PISA 2000) there will be scope for more than one scale but in the minor domains a single scale will be used.

The context questionnaires and their use

To gather contextual information, students and the principals of their schools will also be asked to respond to background questionnaires that will take 20 to 30 minutes to complete. These questionnaires are central tools to allow the analysis of results in terms of a range of student and school characteristics.

The questionnaires will seek information about:

– the students and their family backgrounds, including the economic, social and cultural capital of students and their families;

– aspects of students' lives such as their attitudes to learning, their habits and life inside school and their family environment;

– aspects of schools such as the quality of the school's human and material resources, public and private control and funding, decision-making processes and staffing practices;

– the context of instruction including institutional structures and types, class size and the level of parental involvement.

The first cycle of OECD/PISA will also include an instrument that asks students to report on self-regulated learning. This instrument is based on the following components:

– strategies of self-regulated learning, which govern how deeply and how systematically information is processed;

– motivational preferences and goal orientations, which influence the investment of time and mental energy for learning purposes as well as the choice of learning strategies;

– self-related cognition mechanisms, which regulate the aims and processes of action;

– action control strategies, particularly effort and persistence, which protect the learning process from competing intentions and help to overcome learning difficulties;

– preferences for different types of learning situations, learning styles and social skills required for co-operative learning.

Overall, the OECD/PISA context questionnaires will provide a detailed basis for policy-oriented analysis of the assessment results. Together with information obtained through other OECD instruments and programmes, it will be possible to:

– compare differences in student outcomes with variations in education systems and the instructional context;

– compare differences in student outcomes with differences in curriculum content and pedagogical processes;

– consider relationships between student performance and school factors such as size and resources, as well as the differences between countries in these relationships;

– examine differences between countries in the extent to which schools moderate or increase the effects of individual-level student factors on student achievement;

– consider differences in education systems and the national context that are related to differences in student achievement between countries.

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

OECD/PISA – An evolving instrument

Given the long horizon of the project and the different relative emphases that will be given to the domains in each cycle, the OECD/PISA assessment frameworks clearly represent an instrument that will evolve. The frameworks are designed to be flexible so that they:

– can evolve and adapt to the changing interests of participants; and yet

– give guidance to those enduring elements that are likely to be of continuing concern and therefore should be included in all assessment cycles.

The frameworks will be modified in the light of field trials in the course of 1999 before a final instrument is produced for use in OECD/PISA 2000. Moreover, the development of the survey will continue thereafter, to take account of both changing objectives of education systems and improvements in assessment techniques. The benefits of such development and improvement will, of course, have to be balanced against the need for reliable comparisons over time, so that many of the core elements of OECD/PISA will be maintained over the years.

OECD's objectives are ambitious. For the first time an international assessment of school students aims to determine not just whether they have acquired the knowledge specified in the school curriculum, but whether the knowledge and skills that they have acquired in childhood have prepared them well for adult life. Such outcome measures are needed by countries which want to monitor the adequacy of their education systems in a global context. The ideal will not be instantly achieved, and some of OECD/PISA's goals will initially be constrained by what is practicable in an assessment instrument that needs also to be reliable and comparable across many different cultures. But the objectives are clear, and the assessments that evolve over the coming years will aim to move progressively towards fulfilling them.

> Students in PISA 2000 will have their abilities assessed by being required to perform a set of pencil and paper tasks within a given period of time. The term "test" has varying connotations in different countries, in some cases implying that the results have implications for individual students. PISA's purpose is to survey characteristics of each country's students collectively, rather than to examine individuals. Therefore the term "assessment" is used to describe PISA, even though the conditions experienced by students will be similar to those of a school test.

DEVELOPING OECD/PISA AND ITS ASSESSMENT FRAMEWORKS
A COLLABORATIVE EFFORT

OECD/PISA represents a collaborative effort among the Member governments of the OECD to provide a new kind of assessment of student achievement on a regular basis. The OECD/PISA assessments are jointly developed and agreed by participating countries, and are carried out by national organisations.

A Board of Participating Countries on which each country is represented determines, in the context of OECD objectives, the policy priorities for OECD/PISA and oversees adherence to these priorities during the implementation of the programme. This includes the setting of 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 that are charged with linking the OECD/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;

– the assessment materials have strong measurement properties; and that

– the instruments place an emphasis on authenticity and educational validity.

Through National Project Managers, participating countries implement OECD/PISA at the national level subject to the agreed administration procedures. National Project Managers play a vital role in ensuring that implementation is of high quality, and verify and evaluate the survey results, analyses, reports and publications.

The design and implementation of the surveys, within the framework established by the Board of Participating Countries, is the responsibility of an international consortium led by the Australian Council for Educational Research (ACER). Other partners in this consortium include the Netherlands National Institute for Educational Measurement (CITO), the Service de Pédagogie Expérimentale de l'Université de Liège in Belgium, and WESTAT in the United States.

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 Board of Participating Countries, builds consensus among countries and serves as the interlocutor between the Board of Participating Countries and the international consortium charged with the implementation of the activities. The OECD Secretariat will also produce the indicators, and analyse and prepare the international reports and publications in co-operation with the PISA consortium and in close consultation with Member countries both at the policy level (Board of Participating Countries) and at the level of implementation (National Project Managers).

The OECD/PISA frameworks, described below, have been developed by expert panels under the direction of the ACER. The reading expert panel was chaired by Dr. Irwin Kirsch of Educational Testing Service, the mathematics panel was chaired by Professor Jan de Lange of the University of Utrecht, and the science expert panel was chaired by Professor Wynne Harlen of the Scottish Council for Research in Education. The frameworks were also informed by reviews from expert panels in each of the participating countries. The frameworks were finally adopted by OECD governments, through the Board of Participating Countries.

The development of the OECD/PISA frameworks can be described as a sequence of the following six steps:

– development of a working definition for the domain and description of the assumptions that underlie that definition;
– evaluation of how to organise the set of tasks constructed in order to report to policy makers and researchers on achievement in each assessment domain among 15-year-old students in participating countries;
– identification of a set of key characteristics that should be taken into account when constructing assessment tasks for international use;
– operationalisation of the set of key characteristics that will be 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; and
– preparation of an interpretative scheme for the results.

It is the first four of these steps that are described in this publication. The last two steps will be completed once the results from the field trial assessments are available. The developmental process itself was guided by the following principles:

– a framework should begin with a general definition or statement of purpose – one that guides the rationale for the survey and for what should be measured;
– a framework should identify various task characteristics and indicate how these characteristics will be used in constructing the tasks; and
– variables associated with each task characteristic should be specified, and those that appear to have the largest impact on the variance in task difficulties should be used to create an interpretative scheme for the scale.

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

– 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 evolves, 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;
– linking research, assessment, and public policy promotes not only the continued development and use of the survey, but also the understanding of what it is measuring.

Chapter 1

READING LITERACY

In OECD/PISA, the term *reading literacy* is understood in a broad sense. Since comparatively few young adults in our societies have no literacy skills, the framework does not call for a measure of whether or not 15-year-old students can read in a technical sense. It does reflect, however, contemporary views about reading. These views hold that students, upon leaving secondary school, should be able to construct, extend, and reflect on the meaning of what they have read across a wide range of continuous and non-continuous texts commonly associated with a variety of situations both within and beyond the school doors.

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 concepts of learning, and particularly of lifelong learning, have expanded the perception and requirements of reading literacy. Reading literacy is no longer considered an ability only acquired in childhood during the early school years but is instead viewed as a progressive set of knowledge, skills and strategies which individuals build on throughout life in various contexts and through interaction with their peers.

Cognitive views of reading literacy emphasise the interactive nature of reading and the constructive nature of comprehension (Bruner, 1990; Dole *et al.*, 1991; Binkley and Linnakylä, 1997). The reader generates meaning in response to text by using previous knowledge and a range of textual and situational cues that are often socially and culturally shared. While constructing meaning, the reader uses various processes, skills, and strategies to foster, monitor and maintain understanding. These processes and strategies are expected to vary along with the situation and the purpose as readers interact with a variety of continuous and non-continuous texts.

Two recent international reading literacy assessments (the International Association for the Evaluation of Educational Achievement's Reading Literacy Study – IEA/RLS; and the International Adult Literacy Survey – IALS – undertaken jointly by Statistics Canada and the OECD) have also emphasised the functional nature of reading.

IEA/RLS defined reading literacy as:

"*The ability to understand and use those written language forms required by society and/or valued by the individual.*"

IALS also accentuated the functional nature of reading literacy and particularly its potential in individual and societal development. Its definition focused on information rather than on language forms. Reading literacy was defined as:

"*Using printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential.*"

These definitions of reading literacy focus on the reader's ability to use written or printed texts for purposes required by society or valued by individuals to develop their knowledge and potential. The definitions go beyond simple decoding and literal comprehension and imply that reading literacy incorporates both understanding and use of written information for functional purposes. These definitions, however, do not emphasise the active and initiative role of the reader in understanding or using information. Thus the definition of reading literacy that is used in OECD/PISA is as follows:

> *"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."*[1]

The following remarks further explain this definition.

Reading literacy…

The term "reading literacy" is used in preference to "reading" because it is likely to convey more precisely to a non-expert audience what the survey is measuring. "Reading" is often understood as simply decoding, or reading aloud, whereas the intention of this survey is to measure something broader and deeper. The focus is on the application of reading in a range of situations for various purposes. Historically, the term "literacy" referred to a tool which readers use to acquire knowledge. The term literacy alone is not sufficient since it is too often associated with illiteracy or some minimum level of skill needed to function in a given society. However, the reference to literacy as a tool seems close to the connotations that the term "reading literacy" is intended to convey in OECD/PISA, which covers a broad spectrum of students. Some of these students will go on to a university, possibly to pursue an academic career; some will pursue other secondary or tertiary education in preparation for their participation in the labour force; and some will be entering the workforce directly upon completion of compulsory education. Regardless of their academic or labour force aspirations, students are expected to become active participants in their respective communities.

… is understanding, using, and reflecting on…

The words "reflecting on" were added to "understanding" (from IEA/RLS) and "using" (from IEA/RLS and OECD/IALS) in order to emphasise the notion that reading is interactive: readers draw on their own thoughts and experiences in engaging with a text. Reflection might require readers to think about the content of the text, exploiting their previous knowledge or understanding, or to think about the structure or form of the text.

… written texts…

The words "written texts" are meant to include those texts – printed, hand-written, or displayed electronically – in which language is used. These include visual displays such as diagrams, pictures, maps, and tables or graphs, but do not include film, TV, animated visuals, or pictures without words. These visual texts can occur either independently or be embedded in continuous texts. Written texts also include those in electronic format, even though some of those may be different from written texts in structure and format and may require different reading strategies. It is expected that electronic texts will be used in future survey cycles but will not be included in this cycle because of considerations of time and access. The term "texts" was chosen in preference to the word "information" used in the IALS definition because it was thought that the latter term did not adequately incorporate literature.

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

This phrase is meant to capture the full scope of situations in which reading literacy plays a role, from private to public, from school to work, to lifelong learning and active citizenship. "To achieve one's goals and to develop one's knowledge and potential" spells out the idea that literacy enables the fulfilment of individual aspirations – both defined aspirations such as graduating or getting a job, and those less

1. While the assumptions underlying the definition of reading literacy are commonly understood throughout participating countries, specific words do not exist in some languages. To show that the intended meaning of this section of the document can be translated into other languages without changing the underlying meaning of the term "reading literacy" or the assumptions that have been made about its definition, translations of this section have been developed and are available from the OECD/PISA web site: http://www.pisa.oecd.org.

defined and less immediate which enrich and extend personal life and lifelong education. "Participate" is used in preference to the word "function" used in IALS because it implies that literacy allows people to contribute to society as well as to meet their own needs. The term "function" carries a limiting pragmatic connotation, whereas "participate" includes social, cultural, and political engagement. Participation may include a critical stance, a step towards personal liberation, emancipation, and empowerment. The term "society" includes economic and political as well as social and cultural life (Linnakylä, 1992; Lundberg, 1991, 1997; MacCarthey and Raphael, 1989).

Organisation of the domain ans task characteristics

Having defined the domain of reading literacy and having laid out the set of assumptions that were made in constructing this definition, it is important to set out a framework of how to organise this domain. This organisation needs to focus on how to report the scores that result from administering the pool of reading literacy tasks. This is an important issue because how the domain is organised can affect test design. Research suggests that reading is not a single, one-dimensional skill, and that reading literacy therefore cannot be represented adequately by a single scale or single score along that scale. Determining how many and which scales should be used for reporting reading literacy scores is crucial for ensuring that sufficient numbers of tasks are developed to define and interpret these scales adequately.

Different perspectives can be used to help organise the scales. The easiest way would be to rely on the work of others who have conducted national and international student assessments. Both the United States National Assessment of Educational Progress reading assessment (NAEP) and the IEA/RLS reported scores on three scales. These scales focused on text format. IEA/RLS reported their results for 4th and 9th-grade students on narrative, expository and document scales. The NAEP, using a similar approach, reported proficiencies on three scales: literature – or reading for literary experience; information – or reading to be informed; and documents – or reading to perform a task. These scales are also similar to those used in the International Adult Literacy Survey or IALS. In addition to a quantitative literacy scale, IALS also reported proficiencies along a prose scale and a document scale. In this assessment, prose literacy consisted primarily of expository texts, while for the school-based NAEP surveys there was more balance between narratives and exposition.

A second way to organise the reading tasks is based on the situations from which the tasks are constructed. One of the goals of OECD/PISA is to measure reading literacy not just within an academic setting but across a variety of situations. This is because the assessment seeks to address the question of whether students in the targeted age group are adequately prepared to enter the workforce and to participate as members of their communities.

Another way in which to organise and report internationally on the reading literacy skills would be to use a scheme based on task content; technical content versus humanities is one distinction that has been put forward.

Still another organisational scheme might be based on aspects of reading, such as forming a broad general understanding, retrieving information, developing an interpretation, reflecting on the content of a text, and reflecting on the structure and form of a text. The fine understanding of a text demands that readers engage in all aspects of reading (Langer, 1995) and is in fact a major element in the development of reading literacy tasks in OECD/PISA.

It is believed that the OECD/PISA reading literacy assessment will provide a rich array of data from which a decision will be reached on how to report results most effectively to policy makers, educators, and researchers. This decision will be based on a combination of three criteria: conceptual, empirical, and political. The final decisions about which reporting scales will be used will be made after the data from the OECD/PISA field trial has been collected and analysed.

In addition to organising the reading literacy domain for reporting purposes, it is also necessary to identify task characteristics and to begin to operationalise these characteristics so that the construction and selection of tasks can begin. Only a finite number of characteristics can be manipulated in the construction of tasks, and only a small number of variables associated with these characteristics are likely to play an important role in influencing student performance.

Almond and Mislevy (1998) note that such variables can take on one of five roles in an assessment. They can be used to:

– limit the scope of the assessment;

– characterise features that should be used for constructing tasks;

– control the assembly of tasks into booklets or test forms;

– characterise students' performance in or responses to tasks; or to

– help to characterise aspects of competencies and proficiencies.

Some of these variables can be used to help both in the construction of tasks and in the understanding of competencies, as well as in the characterisation of performance.

A finite number of the task characteristics most relevant for measuring student performance have been selected for constructing and marking tasks. These characteristics are components of the reading process that will be manipulated in the OECD/PISA reading literacy survey to simulate and evaluate the interactive nature of the reading process. They include:

– ***Situation:*** Since adults do not read written texts in a vacuum but within a particular situation, it is important to identify a range of situations from which materials for this reading literacy assessment can be sampled. It should be remembered that one of the goals of OECD/PISA is to move beyond classroom-based texts to include a range of materials that students will also encounter outside their classrooms.

– ***Texts:*** While no one would doubt that a reading literacy assessment should include a variety of material, what is critical to the design and interpretation of the scores that are produced is the range and specific features of the text material which is included in constructing the tasks. Thus, a broad range of both continuous and non-continuous text types are included in OECD/PISA, and consideration will be given to having students read these materials both singly and in combination. That is, they might be asked to read two continuous texts covering a similar topic, or a continuous text and a non-continuous text such as a graph or table.

– ***Test rubric:*** This term refers to the characteristics of the questions and directives in the tasks that are given to the students, the response formats that are used to elicit the responses, and the marking guides that are applied to the responses that students make. Generally speaking, the questions and directives will refer to a goal or purpose which the readers are asked to assume while they are reading and interacting with texts. The reading literacy survey will not rely solely on the use of multiple-choice formats but will include open-ended tasks which will be designed to engage the students in a broader and deeper range of processes and strategies.

In order to use these three main task characteristics in designing the assessment and, later, in interpreting the results, they must be operationalised. That is, the various values that each of these characteristics can take on must be specified. This will allow item developers to categorise the materials that they are working with and the tasks that they construct so that these can then be used to organise the reporting of the data along with the interpretation of results. These variables can also be used to specify what proportions of the assessment ought to come from each category. What role the variables do play in interpretation is, of course, an empirical question, but they cannot play any role if they are not built into the design of the assessment.

Situations

Situation refers more to the uses for which an author composes a text than to location or setting. While it is the intent to assess both the kinds of reading that are associated with school and those that occur beyond the school door, the manner in which the situation is specified cannot simply be based on the place where the reading activity is carried out. For example, textbooks are read both in schools and in homes, and the process and purpose probably differ little from one setting to another. Also, as Hubbard (1989) has shown, some kinds of reading usually associated with out-of-school settings for children, such as rules for clubs and records of games, often take place unofficially at school as well.

Although most reading is a solitary activity, it nonetheless has social aspects. Others are involved in the reading, as authors, as topics, and as those who set tasks (such as teachers). Situation includes reference to the people and (in the case of reading at work) objects that are connected with the reading.

For the purpose of the assessment of 15-year-old students in OECD/PISA, situation can be understood as a categorisation of tasks based on their intended use, on the relations to others implicit or explicit in the task, and on the general contents (see Table 2). Thus, reading a textbook would be an example of an educational situation because its primary use is to acquire information as part of an educational task (use), it is associated with assignments from teachers or other instructors (others), and its content is typically oriented to instruction and learning (content).

Table 2. **Situations for reading**

	Reading for private use	Reading for public use	Reading for work	Reading for education
Others	Self Relatives Friends	Anonymous	Objects Co-workers Managers	Instructors
Use	Curiosity Contact	Information	To do	To learn
Contents	Letters Fiction Biography "How to…" books and magazines Maps	Notices Regulations Programmes Pamphlets Forms	Instructions Manuals Schedules Memos Reports Tables/Graphs	Texts Maps Schematics Tables Graphs

While content is not a variable that is specifically manipulated in this assessment, the sampling of texts is drawn from a variety of situations to maximise the diversity of content that will be included in the reading literacy survey. Close attention is being paid to the origin and content of texts selected for inclusion, as well as to the types of questions and directives that are used to elicit evidence about students' reading literacy. The goal is to reach a balance between constructing tasks that best reflect the broad definition of reading literacy used in OECD/PISA, and recognising the need that these materials should be representative of the linguistic and cultural diversity of participating countries. This diversity will help to ensure that no one group is either advantaged or disadvantaged by the assessment content.

A useful operationalisation of the situation variables can be taken from the Council of Europe's (1996) work on language:

– *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 larger society. This 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 only some 15-year-olds will actually have to read at work, it is important to include tasks that are typical of reading for work in that they are closely tied to the accomplishment of some immediate task and the contents are directly relevant to the goals of this assessment. It is also important to assess the readiness of 15-year-olds to move into the world of work, since many of them will move into the labour force within one to two years. Typical tasks are often referred to as "reading to do" (Sticht, 1975; Stiggins, 1982).

– *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 an instructor. The content is usually designed specifically for the purpose of instruction. Typical tasks are often referred to as "reading to learn" (Sticht, 1975; Stiggins, 1982).

Text types

Reading requires something for the reader to read. In an assessment, that something – a text – must be coherent within itself. That is, the text must be able to stand alone without requiring additional printed material.[2] While it is obvious that there are many different kinds of texts and that any assessment should include a broad range of them, it is not as obvious that there is an ideal categorisation of text types. There are different proposals as to the appropriate categories, many of them created for practical rather than theoretical purposes. All of them share the fact that no particular physical text seems to fit easily into only one category. For example, a chapter in a textbook might include definitions (often identified as a text type), instructions on how to solve particular problems (yet another text type), a brief historical narrative of the discovery of the solution (still another text type), and descriptions of some typical objects involved in the solution (a fourth text type).

It might be thought that a definition, for example, could be extracted and treated as a single text for assessment purposes. But this would remove the definition from the context, create an artificial text type (definitions almost never occur alone, except in dictionaries), and prevent item writers from creating tasks that deal with reading activities which require integrating information from a definition with information from instructions.

Some texts are presented as being accounts of the world as it is (or was) and thus claim to be factual or non-fictional. Fictional accounts bear a more metaphorical relationship to the world as it is, appearing either as accounts of how it might be or of how it seems to be. This distinction is increasingly blurred as authors use formats and structures typical of factual texts in creating their fictions. The OECD/PISA reading assessment will include both factual and fictional texts, and texts that may not be clearly classified as one or the other, but will not attempt to measure differences in reading proficiency between one type and the other.

A more important classification of texts, and one at the heart of the organisation of the OECD/PISA assessment, is the distinction between continuous and non-continuous texts. Continuous texts are typically composed of sentences that are, in turn, arranged in paragraphs. These may fit into even larger structures such as sections, chapters, and books. Non-continuous texts are most frequently organised in matrix format, based on combinations of lists.

Conventionally, continuous texts are formed of sentences arranged in paragraphs. In these texts, organisation is evident in paragraphing, indentation, and the breakdown of text into a hierarchy signalled by headings that help readers to recognise the structure of the text. These markers also provide clues to text boundaries (marking section completion, for example). The finding of information is often facilitated by the use of different font sizes, font types such as italic or bold, and borders or shading. The use of format clues is an essential sub-skill of effective reading.

Organisational information is also signalled by discourse markers. Sequence markers (first, second, third, etc.), for example, signal the relationships between the units which they introduce and indicate how the units relate to the larger surrounding text.

The primary classification of continuous texts is by rhetorical purpose, or text type.

Non-continuous texts, or documents as they are sometimes referred to, can be categorised in two ways. One is the formal structure approach used in the work of Kirsch and Mosenthal.[3] Their work classifies the texts by the way in which the underlying lists are put together to construct the various

2. This does not preclude the use of several texts in a single task, but each of the texts should itself be coherent.
3. The Kirsch and Mosenthal model was set out in detail in a series of monthly columns called "Understanding Documents" published in the *Journal of Reading* between 1989 and 1991.

non-continuous text types. The other uses ordinary descriptions of these texts. The Kirsch and Mosenthal approach is systematic and provides a way of categorising all non-continuous texts, regardless of their use.

Continuous text types

Text types are standard ways of classifying the contents of continuous texts and the author's purpose in such texts. Each type has typical formats in which it occurs. These are noted after each type.[4]

1. *Description* is the type of text in which the information refers to physical, *spatial* properties of objects or characteristics of people. Descriptive texts typically provide an answer to "*what*" questions.
   – *Impressionistic descriptions* present information from the point of view of subjective impressions of relationships, qualities, and spatial directions.
   – *Technical descriptions* present information from the point of view of objective spatial observation. Frequently, technical descriptions use non-continuous text formats such as diagrams and illustrations.
2. *Narration* is the type of text in which the information refers to *temporal* properties of objects. Narration texts typically provide answers to *when*, or *in what sequence* questions.
   – *Narratives* present changes from the point of view of subjective selection and emphasis, recording actions and events from the point of view of subjective impressions in time.
   – *Reports* present changes from the point of view of an objective situational frame, recording actions and events which can be verified by others.
   – *News stories* purport to enable the readers to form their own independent opinion of facts and events without being influenced by the reporter's own views.
3. *Exposition* is the type of text in which the information is presented as composite concepts or mental constructs, or those elements into which concepts or mental constructs can be analysed. The text provides an explanation of how the component elements interrelate in a meaningful whole and often answers *how* questions.
   – *Expository essays* provide a simple explanation of concepts, mental constructs, or conceptions from a subjective point of view.
   – *Definitions* explain how terms or names are interrelated with mental concepts. In showing these interrelations, the definition explains the meaning of "words".
   – *Explications* are a form of analytic exposition used to explain how a mental concept can be linked with words or terms. The concept is treated as a composite whole which can be understood if decomposed into constituent elements and if the interrelations between these are each given a name.
   – *Summaries* are a form of synthetic exposition used to explain and communicate about "texts" in a shorter form than in the original text.
   – *Minutes* are a more or less official record of the results of meetings or presentations.
   – *Text interpretations* are a form of both analytic and synthetic exposition used to explain the abstract concepts which are present in a particular (fictional or non-fictional) text or group of texts.
4. *Argumentation* is the type of text that presents propositions as to the relationships between concepts, or other propositions. Argument texts often answer *why* questions. Another important sub-category of argument texts is persuasive texts.
   – *Comment* relates the concepts of events, objects, and ideas to a private system of thought, values, and beliefs.

4. This section is based on the work of Werlich (1976). It should be noted that the category "hypertext" is not part of Werlich's scheme.

– *Scientific argumentation* relates concepts of events, objects, and ideas to systems of thought and knowledge so that the resulting propositions can be verified.

5. *Instruction* (sometimes referred to as *injunctive*) is the type of text that provides directions on what to do.
   – *Instructions* present directions for certain behaviours in order to complete a task.
   – *Rules, regulations,* and *statutes* specify requirements for certain behaviours based on impersonal authority, such as public authority.
6. *Hypertext* is a set of text slots linked together in such a way that the units can be read in different sequences. These texts frequently have visual supports and may invite non-linear strategies on the part of readers.

Non-continuous texts (structure and format)

Non-continuous texts are organised differently than continuous texts and therefore require different kinds of reading approaches. It is convenient to think about these texts in two ways. The first looks at the principles by which the elements of the text are arranged. This *text structure* variable identifies the features of non-continuous texts that have the same function as the sentence and paragraph features of continuous text. The second approach identifies some common *formats* for these texts.

• Non-continuous texts by structure

All non-continuous texts can be shown to be composed of a number of lists. Some are just simple lists, but most are comprised of combinations of several lists. This analysis of these texts does not refer to their use or employ the common labels often attached to non-continuous texts, but does identify key structural features that are common to a number of different texts. A complete description of any non-continuous text requires both a structural and a format category. Readers who understand the structure of texts are better able to identify the relationships between the elements and to understand which texts are similar and which are different.

1. *Simple lists* contain only a single collection of elements. The list of books to be read in a literature course is an example of a simple list, as is the list of students on the honours list. The elements in the list may be ordered, as when the list of students in a class is arranged alphabetically by last name, or unordered, as in the list of supplies to buy for an art class. It is easier to find items in the former than in the latter. If the unordered list is long, it may be difficult to determine whether an item is listed or not. This should easily be possible on the ordered list, provided that one knows the ordering principle.
2. *Combined lists* are made up of two or more simple lists in which each element in one list is paired with an element in another list. One of the lists may be taken as the primary list (indexing list); this primary list is ordered to make it easier to find items in it, so that the parallel information in the other lists can also be found. An elementary combined list might be a list of students with a corresponding list of marks in an assessment. Items may occur more than once in one of the lists, though this seldom happens with the primary list. For example, in a list of students and marks, any mark may appear several times. A combined list may have many component lists as in a list of popular songs which may have the song title, the singer(s), the record label, and the number of weeks it has been on the hit chart. Searches of the non-indexing list are more difficult, and it may be difficult to know whether all relevant information has been obtained. Thus, using the list of students and marks to find out what mark a particular student received will be easy, especially if the students' names are in alphabetical order. It may be more difficult to find all the students who did not receive a pass mark in the assessment.
3. *Intersecting lists* consist of three lists which are not parallel, but intersecting, and form a row and column matrix. The typical intersecting list is a television schedule that consists of a list of times, a list of channels, and a list of programmes. The programmes occur in the cells at the intersection of a time (usually defining the columns) and a channel (usually defining the rows). In academic

settings a department may prepare a table of courses in a matrix format, with the columns representing days; the rows, times; and the cell entries, the course(s) offered at a particular time on a particular day. This makes it easy for students to identify courses that do not conflict in time. In an intersecting list the cell entries are all of a single kind (course titles, TV programmes, etc.). Many statistical tables are intersecting lists. For example, a table that lists the unemployment rates for large cities is likely to have the cities as rows, columns as particular dates, and cell entries as the actual rates for the cities at those times. The table may be designed to permit comparisons between dates, as when there are several columns, each representing a different period (months, years, etc.).

4. *Nested lists* consist of a set of combined lists. For example, in some intersecting lists the column categories, such as days of the week, intersect not only with the row categories (times) but also with a fourth list, such as departments in a university. For a true nested list, the same type of category must be used in each of the intersecting lists. The intersecting list of unemployment rates may have separate entries under each month for males and females; in this case, gender is nested under month.

5. *Combination lists* are those in which several types of lists, or several lists of the same type, are joined into one list. For example, the intersecting list created by the statistical table of unemployment rates in different months for large cities may be combined with another intersecting list of month-to-month changes in the unemployment rates for those cities.

• Non-continuous texts by format

Classifying non-continuous texts by their formats provides another perspective on these texts. Every non-continuous text can be classified by both structure and format. For example, forms are one of the format categories, but every form also has a structure, most commonly a combined list in which a list of labels is paired with a list of blanks to be filled in with information that corresponds to the labels. A timetable (for buses, railways, or airlines) is a table format whose structure is often an intersecting or nested list. Recognising the format is important because texts with the same structure may be laid out on the page quite differently. For example, a table of contents for a book and a form are usually combined lists. In a form the two lists are the label and the blank, as noted above. In a table of contents the two lists are the chapter titles and the pages on which the chapter starts; these are paired just as the label and blank field in a form are. But no one would confuse a form with a table of contents.

1. *Forms* are structured and formated 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.

2. *Information sheets*, as opposed to forms, provide rather than solicit information. They summarise information in a structured way and in such a format that the reader can easily and quickly find specific pieces of information. Information sheets may contain various text forms, together with lists, tables, figures, and sophisticated text graphics (headings, fonts, indentation, borders, etc.) which summarise and highlight information for the eye. Timetables, price lists, catalogues and programmes are examples of this form of document.

3. *Vouchers* testify that their owner is entitled to certain services. The information which they contain must be sufficient to show whether the voucher is valid or not. Typical examples are tickets, invoices, etc.

4. *Certificates* are written acknowledgements of the validity of an agreement or a contract. They are formalised in content rather than format. They usually require the signature of one or more persons authorised and competent to give testimony of the truth of the given statement. Warranties, school certificates, diplomas, contracts, etc., are documents that have these properties.

5. *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 format.
6. *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.
7. *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.
8. *Tables and matrices.* 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 timetables, spreadsheets, order forms, and indexes.
9. *Lists* are the most elementary non-continuous texts. They consist of a number of entries that share some property(ies), which may be used as a label or title for the list. Lists may have their entries ordered (*e.g.* the alphabetically arranged names of students in a class) or unordered (*e.g.* a list of goods to be bought in a shop).
10. *Maps* are non-continuous texts that indicate the geographical relationships between places. There is a variety of types of maps. Road maps mark the distances and routes between identified places. Thematic maps indicate the relationships between locations and social or physical features.

Test rubrics and task characteristics

There are three sets of variables that make up the test rubrics: questions or directives, which set out the task for the student; response formats, which set out the ways in which students are requested to demonstrate their proficiency at the task; and rules for marking, which specify how students' responses are to be evaluated. Each of these will be discussed in turn, though the first requires considerably more attention.

Questions and directives

It is possible to look at the tasks set out for students from a macro perspective and from a micro perspective. At the macro level, the tasks can be identified as requiring one or more of five broad aspects of reading. At the micro level, several variables have been shown to play a role in determining the difficulty of tasks along a particular scale.

- Macro aspects

Because it is assumed that the majority of 15-year-olds in the participating countries will have mastered the basic decoding of texts, it is not necessary to focus explicitly on the underlying skills in reading. Rather, the assessment will focus on more complex reading literacy strategies (Dole *et al.*, 1991; Paris, Wasik and Turner, 1991).

In an effort to provide for authentic reading situations, the OECD/PISA reading assessment will measure the following five aspects associated with the full understanding of a text, whether the text is continuous or non-continuous. Students are expected to demonstrate their proficiency in all these aspects:

– forming a broad general understanding;

– retrieving information;

– developing an interpretation;

– reflecting on the content of a text; and

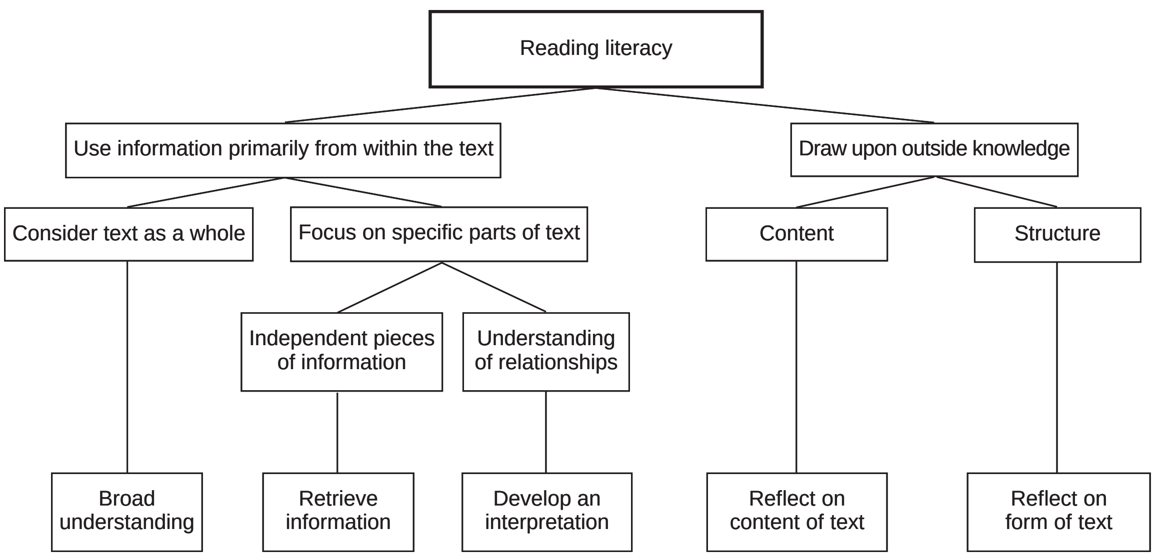
– reflecting on the form of a text.

The complete understanding of a text involves all of these aspects. It is expected that all readers, irrespective of their overall proficiency, will be able to demonstrate some level of competency with each of them (Langer, 1995). While the five aspects are interrelated – each may require many of the same underlying skills – successfully accomplishing one may not be dependent upon successful completion of any other. They are seen by some as being in the repertoire of each reader at every developmental level rather than forming a sequential hierarchy or set of skills. Of course, it will be possible, and necessary, to investigate this assumption once the items have been developed and the field-test data have been collected.

These five aspects of reading will be operationalised through a set of questions and directives presented to students taking the OECD/PISA assessment. Through their responses to these questions and directives, students will be able to provide evidence that they can understand, use, and reflect on a range of continuous and non-continuous texts.

Figure 1 identifies the key distinguishing characteristics of the five aspects of reading being measured for OECD/PISA. While this figure necessarily oversimplifies each aspect, it provides a useful scheme for organising and remembering the relationships between them.

Figure 1. **Characteristics distinguishing the five aspects of reading**



The five aspects 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 more heavily upon outside knowledge.

– A second distinguishing characteristic involves the extent to which the reader is asked to consider the text as a whole or to focus on specific pieces of information contained within the text.

– Sometimes readers are expected to retrieve specific or independent pieces of information while at other times they are directed to demonstrate their understanding of the relationships between parts of the text. Focusing on isolated or independent pieces instead of relationships between parts of the text is the third distinguishing characteristic.

– The fourth characteristic marks whether the reader is directed to deal with the content or substance of the text rather than its form or structure.

The five aspects of reading are represented in the last line of Figure 1, at the ends of the various branches. Starting at the top of the figure and following each branch one can see which characteristics are associated with each aspect.

While Figure 1 necessarily simplifies the content and complexity of each aspect, an initial attempt is made here to define these aspects of understanding a text operationally, and to associate them with particular kinds of questions and directives. Although each aspect is discussed in terms of a single text, it should be understood that they also apply to multiple texts when these form a unit.[5] The description of each aspect has two parts. The first provides a general overview of the aspect, while the second describes particular ways in which the aspect might be assessed.

a) Forming a broad general understanding

Often a reader wishes to obtain just a broad understanding of the text before choosing to read in more detail. Through this initial reading, mature readers may determine whether the text, either continuous or non-continuous, suits their intended goal.

To form a broad general understanding of the text, a reader must consider it as a whole or in a broad perspective. This resembles the first encounter with a person or place in that the reader makes hypotheses or predictions of what the text is about based on first impressions. These impressions are quite general but very important for selecting the most relevant and interesting reading material.

In that the tasks asking for a broad understanding are based on the text itself they are similar to those tasks that require a reader to retrieve information or develop an interpretation. But, in contrast to these other tasks, in order to form a general understanding, the reader must grasp the essence of the text as a whole – explaining what it is about, identifying the topic under discussion, etc. Several components are important, such as determining the main idea or the topic, and identifying the general use of the non-continuous text.

There are various assessment tasks in which readers are asked to form a broad general understanding. Students may demonstrate initial understanding through identifying the main topic or message or through identifying the general purpose or use of the text. Examples of such tasks are those that ask 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 which direct the student to describe the main character, setting or milieu of a story; to identify a theme or message of a literary text; to explain the purpose or use of a map or a figure; to identify the main topic or audience of an e-mail message; to work out what kind of book could contain certain types of text; or to grasp a general view or purpose of a home page on the Internet.

Some of the tasks seeking a general understanding might require making a match between one piece of text and the question. For example, this would happen when a theme or main idea is explicitly stated in the text. Other tasks might require the student to focus on more than one specific reference in the text – for instance, if the reader has to deduce the topic 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 ones. 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.

b) Retrieving information

In the course of daily life, readers often need a particular piece of information. They may need to look up a telephone number. They may want to check the departure time for a bus or train. They may want

5. It is intended that exemplary items will be used to illustrate each of these aspects. Inclusion in this public document would jeopardise the secure nature of the assessment, however. Examples that help to define and distinguish each aspect are expected to appear after field-trial data have been collected and analysed and the final items have been selected.

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 and search the text, and locate and select relevant information. The processing involved in this aspect of reading is most frequently at the sentence level, though in some cases the information may be in two or more sentences or in different paragraphs.

Successful mastery of retrieving information requires immediate understanding. Finding the needed piece of information may require that the reader processes more than one piece of information. For example, in determining which bus will allow him or her to leave at the latest hour and still reach the intended destination on time, a reader might refer to a bus timetable and compare the arrival and departure times of various buses which cover the route. To do this the reader would necessarily search for more than one piece of information.

In assessment tasks that call for the retrieval of information, students must match information given in the question with either literal or synonymous information in the text and use this to arrive at the new information requested. In this, the retrieval of information is based on the text itself and on explicit information included in it. Retrieval tasks require the student to find information based on conditions or features specified in the questions or directives. The student has to detect or identify the essential elements of a message: characters, pace/time, setting, etc., and then search for a match that may be literal or synonymous.

Retrieval tasks can also involve coping with some degree of ambiguity. For example, the student may be required to select explicit information, such as an indication of the time or place in a text or in a table. A more difficult version of the same type of a task might involve finding synonymous information. This sometimes relies on categorisation, or it may require discriminating between two similar pieces of information. By systematically varying the elements that contribute to difficulty, measurement of various levels of proficient performance associated with this aspect of comprehension can be achieved.

c) Developing an interpretation

Developing an interpretation requires readers to extend their initial impressions so that they reach a more specific or complete understanding of what they have read. This involves going through the text and linking up information between its various parts, as well as focusing on specific details as parts of the whole.

Tasks in this category call for logical understanding: the reader must process the arrangement of information in the text. To do so, the reader must understand the interaction between local and global cohesion within the text. 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. In more difficult instances (*e.g.* to indicate relations of cause and effect), there might not be any explicit markings.

A text contains more information than what is explicitly expressed. Drawing inferences is an important mental operation because they serve a variety of functions in text comprehension. The inferences make use of information and ideas activated during reading yet not explicitly stated in the text. They depend (more or less) on knowledge about the world brought to bear by the reader. Some are considered necessary for comprehension and are related to the processing of linguistic devices (for example, reference chains); they play a strong part in the coherence of the interpretation built in the course of reading. Others create new information based on the data contained in the text and in the reader's knowledge.

Examples of tasks that might be used to assess this aspect include comparing and contrasting information, drawing inferences, and identifying and listing supporting evidence. "Compare and contrast" tasks require the student to integrate two or more pieces of information from the text. To process either explicit or implicit information from one or more sources in compare and contrast tasks, the reader must often infer an intended relationship or category. Tasks that ask the student to make inferences about the author's intent, and to identify the evidence used to infer that intent, are also examples of tasks that test this aspect of comprehension.

Further examples of questions characteristic of this aspect include inferring meaning from context, identifying a particular character's motive or intention and identifying cause and its effect.

d) Reflecting on the content of a text

Reflecting on the content of a text requires that the reader connects information found 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. In many situations, readers must know how to justify and maintain their own point of view. To do so, readers must be able to develop an understanding of what is said and intended in a text, and must 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, and the ability to reason abstractly.

This aspect of comprehension requires a high level of metacognitive ability. Readers must monitor their own thinking and reaction to a text while testing potential mental models. To meet the demands of this type of task, relevant information must be called on and arranged in a coherent fashion.

Assessment tasks representative of this category of processing would 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.

e) Reflecting on 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. These tasks include critical evaluation, and appreciation of the impact of such textual features as irony, humour and logical organisation. This aspect includes the ability to ferret out bias and to recognise instances of subtle persuasive nuance.

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 characteristics 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. In-depth processing of this nature calls for such activities as reasoning, critical analysis, explanation of whether the author conveys meaning adequately, distinguishing of fact from opinion, etc. The reader is expected to select important units in the text, to integrate secondary units and to argue a position.

Some examples of assessment tasks characteristic of reflecting on 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 identify or comment on the author's use of style and what the author's purpose and attitude are.

• Micro aspects

In applying the five aspects in which students will be asked to demonstrate their proficiency, three process variables can be considered that are drawn from reading and literacy research conducted in other international surveys (IEA/RLS and IALS). These variables are: the type of information requested, the type of match between the information given and that requested, and the plausibility of distracting information. The paragraphs that follow set out the general properties of each of these three characteristics, and considerations affecting the format in which questions are to be answered, and how these answers are to be marked.

a) Type of information requested

This refers to the kinds of information that readers identify to answer an assessment question successfully. The more concrete the information requested, the easier the task is judged to be. In earlier research based on large-scale assessments of adults' and children's literacy (Kirsch, 1995; Kirsch and Mosenthal, 1994; Kirsch, Jungeblut and Mosenthal, 1998), the variable relating to the type of information was scored on a 5-point scale. A score of 1 represented information that was the most concrete and therefore the easiest to process, while a score of 5 represented information that was the most abstract and therefore the most difficult to process. For instance, questions which asked students to identify a person, animal, or thing (*i.e.* imaginable nouns) were said to request highly concrete information and were assigned a value of 1. Questions asking respondents to identify goals, conditions or purposes were said to request more abstract types of information. Such tasks were judged to be more difficult and received a value of 3. Questions that required students to identify an "equivalent" were judged to be the most abstract and were assigned a value of 5. In such cases, the equivalent tended to be an unfamiliar term or phrase for which respondents had to infer a definition or interpretation from the text.

b) Type of match

This refers to the way in which students process text to respond correctly to a question. It includes the processes used to relate information in the question (the given information) to the necessary information in the text (the new information) as well as the processes needed to either identify or construct the correct response from the information available.

Four types of matching strategies were identified: locating, cycling, integrating and generating. Locating tasks require students to match one or more features of information stated in the question to either identical or synonymous information provided in the text. Cycling tasks also require students to match one or more features of information, but unlike locating tasks, they require respondents to engage in a series of feature matches to satisfy conditions stated in the question. Integrating tasks require students to pull together two or more pieces of information from the text according to some type of specified relationship. For example, this relationship might require students to identify similarities (*i.e.* make a comparison), differences (*i.e.* contrast), degree (*i.e.* smaller or larger), or cause-and-effect relationships. This information may be located within a single paragraph or it may appear in different paragraphs or sections of the text. In integrating information, students draw upon information categories provided in a question to locate the corresponding information in the text. They then interrelate the textual information associated with these different categories in accordance with the relationship specified in the question. In some cases, however, students must generate these categories and/or relationships before integrating the information stated in the text.

In addition to requiring students to apply one of these four strategies, the type of match between a question and the text is influenced by several other processing conditions which contribute to a task's overall difficulty. The first of these is the number of phrases that must be used in the search. Task difficulty increases with the amount of information in the question for which the student must search in the text. For instance, questions that consist of only one main clause tend to be easier, on average, than those that contain several main or dependent clauses. Difficulty also increases with the number of responses that students are asked to provide. Questions that request a single answer are easier than those that require three or more answers. Further, questions which specify the number of responses tend to be easier than those that do not. For example, a question which states, "List the three reasons..." is easier to answer than one which says, "List the reasons...". Tasks are also influenced by the degree to which students have to make inferences to match the given information in a question to corresponding information in the text, and to identify the information requested.

c) Plausibility of distractors

This concerns the extent to which information in the text shares one or more features with the information requested in the question but does not fully satisfy what has been requested. Tasks are judged to be easiest when no distractor information is present in the text. They tend to become more difficult as the number of distractors increases, as the distractors share more features with the correct response, and as the distractors appear in closer proximity to the correct response. For instance, tasks tend to be judged more difficult when one or more distractors meet some but not all of the conditions specified in the question and appear in a paragraph or section of text other than the one containing the correct answer. Tasks are judged to be most difficult when two or more distractors share most of the features with the correct response and appear in the same paragraph or node of information as the correct response.

d) Response formats

Both multiple-choice and constructed-response questions have been used in assessing reading proficiency, but the testing literature provides little guidance as to what strategies or processes are best measured by what formats. As Bennett (1993) noted, "Despite the strong assertions by cognitive theorists, the empirical research has afforded only equivocal evidence that constructed-response tasks necessarily measure skills fundamentally different from the ones tapped by multiple-choice questions" (p. 8). In particular, Traub's survey of research on the differences between the two response formats in reading comprehension tests concluded that there was no sign of a strong format effect (Traub, 1993).

The empirical literature on format effects, however, is quite limited. Traub's survey found only two studies, one with college students and one with students in the third grade. Significantly, though, the one with college students (Ward, Dupree and Carlson, 1987) did measure the more complex aspects of comprehension. However, in his presidential address to the American Psychological Association, Frederickson (1984) noted that the real test bias stems from limitations imposed by the sole use of multiple-choice items. In addition, students in some OECD countries may not be familiar with the format of standardised multiple-choice items. Therefore, including a mix of open-ended items will certainly provide a better balance of the types of tasks with which students in classrooms around the world are familiar. This balance may also serve to broaden the constructs of what is being measured.

There is a great range of constructed-response tasks. Some require little judgement on the marker's part; these include tasks that ask the reader simply to mark parts of the text to indicate an answer or to list a few words. Others require considerable subjective judgement by markers, as when the reader is asked to summarise a text in his or her own words.

Given the lack of strong evidence of a method effect, and advice from item developers, it seems wisest to include both multiple-choice and constructed-response items in the reading literacy assessment.

e) Marking

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

Marking is relatively simple with dichotomous constructed-response items, but the specification of correct answers is more difficult. The more students are expected to generate ideas rather than just to identify information in the text, the greater will be the differences among correct answers. Considerable training and monitoring of markers will be required to ensure comparability from marker to marker, even within one country. A balance needs to be found between specificity and openness. If marking guidelines are too specific, then oddly phrased correct answers may be marked as incorrect; if they are too open, then responses and answers that do not fully satisfy the task may be marked as correct.

Constructed-response items lend themselves especially to partial credit scoring, though this does add some complexity to the marking process (and to the development of marking guidelines). Partial credit marking also enables the use of a variety of tasks in which one type of response indicates a more complex understanding of the text than another response, yet both are "correct" responses. It is recommended that partial credit marking be used, at least for the more complex constructed-response items.

Assessment structure

In this section, the distribution of the reading literacy assessment tasks between the various situations, text formats, aspects and item types is described.

One obvious way to distribute the reading literacy tasks in the assessment is to do so evenly across the four situations (Table 3). However, the occupational situation will be given less weight for two reasons. First, it is important to reduce the potential dependence on specific occupational knowledge that can result when occupational texts are selected. Second, it is expected that the same types of questions and directives can be constructed from the other situations, in which 15-year-old students may have better access to the content.

Table 3. **Recommended distribution of reading tasks by situation**

Situation	% of total tasks
Personal	28
Educational	28
Occupational	16
Public	28
Total	**100**

The distribution and variety of texts that students are asked to read for OECD/PISA is an important characteristic of the assessment. Tables 4 and 5 show the recommended distributions of continuous and non-continuous texts. It can readily be seen that continuous texts are expected to represent about two-thirds of the texts contained in the assessment. Within this category, the largest percentage should come from expository materials (33 per cent) while the smallest percentage should represent injunctive texts (7 per cent). The remaining types of continuous texts should be evenly distributed at about 20 per cent each. Non-continuous texts are expected to represent about one-third of the texts in the reading literacy assessment. The overwhelming majority (66 per cent) will be either tables or charts and graphs. The remaining non-continuous texts will be maps, advertisements, and forms that 15-year-olds are expected to be able to read and use. It is important to keep in mind that these percentages are targets for the main assessment and not for the field trial. The selection of texts for the field trial and then for the main assessment will not be determined solely on structural characteristics, such as format and text type. Consideration will also be given to cultural diversity, range of difficulty across texts, potential interest to students and authenticity.

Table 4. **Recommended distribution of reading tasks by continuous text type**

Text type	% of continuous texts	% of test
Narrative	20	13
Expository	33	22
Descriptive	20	13
Argumentative/Persuasive	20	13
Injunctive	7	5
Total	**100**	**66**

Table 6 shows the recommended distribution of reading literacy tasks by each of the five aspects defined above. The largest single percentage of tasks represents developing an interpretation with slightly more than two-thirds of the tasks covering the first three aspects (70 per cent). Each of these three

Table 5. **Recommended distribution of reading tasks by non-continuous text type**

Text type	% of non-continuous texts	% of test
Charts/Graphs	33	11
Tables	33	11
Schematics	10	3
Maps	10	3
Forms	8	3
Advertisements	6	2
Total	**100**	**33**

Table 6. **Recommended distribution of reading tasks by aspects of reading**

Aspect	% of test
Retrieving information	20
Broad understanding	20
Developing an interpretation	30
Reflecting on content	15
Reflecting on form	15
Total	**100**

Table 7. **Recommended distribution of tasks by text format and aspect**

Aspect	% of test	% of continuous texts	% of non-continuous texts
Retrieving information	20	13	7
Broad understanding	20	13	7
Developing an interpretation	30	20	10
Reflecting on content	15	10	5
Reflecting on form	15	10	5
Total	**100**	**66**	**34**

aspects – broad understanding, retrieving information and developing an interpretation – focuses on the degree to which the reader can understand and use information contained primarily within the text. The remaining tasks (30 per cent) will require the student to reflect on either the content or information provided in the text or on the structure and form of the text itself. Table 7 provides the distribution of tasks by text format and aspect.

In determining just what proportion should be constructed-response items, it is necessary to make some assumptions about the distribution of tasks both on a practical and on a conceptual basis. Table 8 shows the proposed distribution of constructed-response and multiple-choice tasks by the five aspects of reading.

Table 8 indicates that approximately 45 per cent of the reading literacy assessment will be constructed-response items that require judgement on the part of the marker. The other 55 per cent will consist of multiple-choice items and those constructed-response items that require little subjective judgement on the part of the marker. This table also reveals that while multiple-choice and constructed-response items will cut across the five aspects, they will not be distributed evenly. That is, a larger percentage of multiple-choice items will be associated with the first three aspects of reading.

Table 8. **Recommended distribution of constructed-response and multiple-choice tasks by the five aspects of reading**

Aspect	% of test	% of *tasks* requiring constructed response	% of *test items* requiring constructed response	% of *test items* requiring multiple choice
Retrieving information	20	35	7	13
Broad understanding	20	35	7	13
Developing an interpretation	30	35	11	19
Reflecting on content	15	65	10	5
Reflecting on form	15	65	10	5
Total	**100**		**45**	**55**

Reporting scales

To meet the aims of OECD/PISA, the development of scales describing student achievement is essential. The process of developing the reporting scales will be iterative, in that initial proposals that are based on past experience and research in the field of learning and cognitive development in mathematics are further developed on the basis of empirical evidence collected during the OECD/PISA field trial.

The two organising characteristics of the reading framework that are being considered for use as reporting scales are the text types (continuous and non-continuous) and the macro aspects (forming a broad general understanding, retrieving information, developing an interpretation, reflecting on the content of a text, and reflecting on the form of a text). This means that there will be either two or five reading reporting scales, in addition to the overall reading literacy scale.

Other issues

Several other issues need to be addressed, in addition to how reading literacy will be defined and measured in this assessment. Three of these areas overlap with the context questionnaires that students will answer and two others concern relations with other assessments. These issues are elaborated in the following sections.

Questionnaire issues

Several issues are most appropriately addressed through questionnaire items rather than through an assessment. One is the collection of information on reading practices and interests, another covers some aspects of metacognition, and a third has to do with the role of technology in the lives of students participating in OECD/PISA.

Reading practices and interests

The OECD/PISA student context questionnaire will include a set of questions designed to assess students' reading practices and to explore the context of their reading activities in general, both in and out of school, as was the case with other large-scale surveys of reading and literacy (Council of Ministers of Education, Canada, 1994; Elley, 1992; Jones, 1995; Smith, 1996; Taube and Mejding, 1997). Such questions are of both descriptive and explanatory interest. That is, they will serve to characterise the populations of 15-year-old students in terms of their access to printed materials, their interest in and attitude toward various literacy activities, and their actual practices. In addition, these elements may help to explain some of the variation in reading literacy proficiency observed among 15-year-old students participating in OECD/PISA.

Two main principles have guided the selection of what types of information to include in the survey:

– the types of information assessed must be relevant in terms of educational policy;

– the reading assessment should have its counterpart in the questionnaires, seeking information about students' attitudes towards reading and reading practices.

This portion of the contextual questionnaire will include:

– The *degree of exposure to various kinds of print at home, at school, or in a public environment*. This includes the following questions: number of books at home, whether the student owns personal books, purchase of a daily newspaper or weekly magazine on a regular basis in the family, visits to a public or school library, etc.

– *Reading practices/habits*. It is important to ensure that the variety and frequency of the different types of print are covered in relation to the various types of text and text formats distinguished in the reading framework, and in relation to their assessment context. Considering the time limitations and the methodological problems brought about by a reading activities inventory, a reasonable balance must be found between the need to list different sets of print in various types of situations on the one hand (in order to assess variety), and the various constraints on the other.

– *Attitudes toward reading and reading interests*. Attitudes toward reading, and motivation, are likely to influence reading practices and achievement; action can also be taken with regard to these aspects by the creation of a favourable climate for reading literacy in and out of school. In OECD/PISA, this aspect is assessed by means of a number of targeted questions that require little response time (*e.g.* Do you like receiving books as presents?, Do you enjoy visiting a library? etc.); preference for reading is also compared with that for other leisure activities (TV, music, going out, video games, etc.). The answers to these types of questions could prove to be less dependent on the compliance effects that are frequently observed in the assessment of reading practices.

Metacognition

A number of studies have shown a relationship between metacognitive knowledge and achievement in young readers (Ehrlich *et al.*, 1993; Ehrlich, 1996). Differences have been found in readers' understanding of the concept of reading, in the goals and purposes for reading, in their strategies for dealing with comprehension, and in the detection of inconsistencies. These differences occur in terms of two basic components of metacognition: students' knowledge about cognition and regulation of cognition. The first component concerns the ability to reflect on our own cognitive processes, and includes knowledge about when, how, and why to engage in various cognitive activities. The second, regulation, concerns the use of strategies that enable us to control our cognitive efforts (Baker, 1991).

Interest in measuring metacognition as part of OECD/PISA rests on the belief that results can yield information relevant to policy makers and can influence the practice of reading and learning, especially since it is believed that these skills can be taught and can be applied more generally than merely in reading tasks. The challenge rests in trying to devise a way of measuring metacognition in large-scale surveys such as OECD/PISA. Most of the literature is based on experimental studies completed with students younger than those participating in OECD/PISA (Myers and Paris, 1978). As a result, there appears to be no existing instrument that can be assumed to yield reliable and valid results.

Since no reliable instrument exists for measuring metacognition among 15-year-old students, and adequate resources are not available to design and build such an instrument, a decision was made that metacognition would not be part of the reading literacy field trial or main assessment in the first cycle. Given the interest in this topic, the development of a metacognition instrument is to be considered for future cycles.

Technology

In our rapidly changing world, the investigation of reading habits and metacognition needs to be extended to include questions relating to technology, especially the computer. Clearly, the availability of electronic texts and their use to obtain and exchange information will become more and more important in students' lives in coming years.

To prepare for a greater involvement of technology in future cycles, the OECD/PISA survey will include a short questionnaire to gather information about students' access to a computer either at home, at school, at work, or in their community; attitudes towards using a computer; frequency of computer use in various settings; and types of activities they engage in with computers.

Links to other assessment areas

It is important to make use of as much information as is available to provide detailed interpretations of the achievement data from the assessment. Some additional information will come from the questionnaire data, but relations with other assessments will also provide data which may enrich the interpretation of the results.

OECD/PISA is not the only assessment of reading proficiency. As mentioned before, the International Adult Literacy Survey (IALS) assessed adult proficiency and it would be useful if OECD/PISA results could be interpreted in the context of this study as well, despite significant differences in the nature and design of the OECD/PISA and IALS instruments. Links with IALS would also provide a way of connecting the student measures with adult measures. This can be achieved in part by including assessment tasks from IALS to provide a direct connection. Thought has been given to the number of items from IALS that would need to be included in the OECD/PISA reading assessment to provide both a conceptual and a statistical link. In addition, trade-offs between including sufficient items to provide these links and wanting to measure something that extends beyond these assessments have been considered.

It would not be possible to form a link with more than one of the three IALS scales, as there would not be time in the assessment to include such a range of IALS items. As it seems likely that the OECD/PISA assessment will have fewer tasks based on non-continuous than on continuous texts, it would be unwise to use the IALS document scale, as the number of IALS tasks necessary for a link might well dominate this category of OECD/PISA. Because the IALS quantitative scale is based mostly on non-continuous tasks, the same objection would apply to it. IALS prose literacy tasks do seem appropriate for use in OECD/PISA, however, because they would not dominate the continuous texts.

To explore this link, two blocks of items from the IALS prose scale have been included in the field trial for OECD/PISA in order to determine how well the item parameters fit the OECD/PISA populations. Assuming that most of the parameters fit, an adequate number of IALS prose literacy tasks will be included in the main assessment for OECD/PISA. These items will serve a dual purpose. First, the IALS item parameters will be used to estimate the prose literacy proficiency of the various OECD/PISA countries. Second, these items will be coded as part of the OECD/PISA reading literacy framework and used to help estimate the reading literacy proficiencies of 15-year-old students in participating countries.

As OECD/PISA is assessing other achievement areas (mathematics and science), though less intensively, there is an opportunity to develop estimates for the relationships between these areas. Simple approaches would only use the fact that similar populations in the same countries are being assessed at the same time, and would not attempt any modelling of these relationships. More interesting, but more complex, approaches would have items that cross assessments so that the relationship between reading and science or mathematics could be directly modelled by using the results from the common items. Psychometric models for such cross-assessments do exist, but the test tasks require careful development and the interpretation of such conditional results is much less straightforward than that of traditional models, in which each task receives a unique position on a single scale. Because the information gained from cross-item approaches is so important, the OECD/PISA field trial will include a block of fully integrated items (items that are marked for more than one assessment area) as well as several blocks that contain distinct reading, science, and mathematics items that are based on common texts or situations.

Chapter 2

MATHEMATICAL LITERACY

Definition of the domain

The mathematical literacy domain is concerned with the capacity of students to draw upon their mathematical competencies to meet the challenges of the future. It is concerned with students' capacities to analyse, reason, and communicate ideas effectively by posing, formulating and solving mathematical problems in a variety of domains and situations.

The definition of mathematical literacy for OECD/PISA is:

"*Mathematical literacy is an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded mathematical judgements and to engage in mathematics, in ways that meet the needs of that individual's current and future life as a constructive, concerned and reflective citizen.*"

Some explanatory remarks are required to explain this definition of the domain.

Mathematical literacy…

The term *literacy* has been chosen to emphasise that mathematical knowledge and skills as defined within the traditional school mathematics curriculum do not constitute the primary focus of OECD/PISA. Instead, the emphasis is on mathematical knowledge put to functional use in a multitude of different contexts and a variety of ways that call for reflection and insight. Of course, for such use to be possible and viable, a great deal of fundamental mathematical knowledge and skills (as often taught in schools) are needed. In the linguistic sense, reading literacy cannot be reduced to, but presupposes, a wide vocabulary and a substantial knowledge of grammatical rules, phonetics, orthography, and so forth. In the same way, mathematical literacy cannot be reduced to, but presupposes, knowledge of mathematical terminology, facts, and procedures, as well as skills in performing certain operations, and carrying out certain methods.

... the world…

The term *the world* means the natural, social and cultural setting in which the individual lives. As Freudenthal (1983) stated: "Our mathematical concepts, structures and ideas have been invented as tools to organise the phenomena of the physical, social and mental world."

... to engage in…

The term *engage in* is not meant to cover only physical or social acts in a narrow sense. The term includes also communicating, taking positions towards, relating to, assessing and even appreciating mathematics. Thus the definition should not be seen to be limited to the functional use of mathematics. The aesthetic and recreational elements of mathematics are encompassed within the definition of mathematical literacy.

... current and future life

The phrase *an individual's current and future life* includes his or her private life, occupational life, and social life with peers and relatives as well as his or her life as a citizen of a community.

A crucial capacity implied by this notion of mathematical literacy is the capacity to pose, formulate and solve mathematical problems within a variety of domains and situations. The situations range from the purely mathematical problems to those in which no mathematical structure is obvious at the outset – *i.e.* where the mathematical structure must first be identified by the problem poser or solver.

It is also of importance to emphasise that the definition is not just concerned with knowing mathematics at some minimal level, but also with using mathematics in a whole range of situations.

Attitudes and emotions, such as self-confidence, curiosity, a feeling of interest and relevance, and a desire to do or understand things, to name but a few, are not components of the OECD/PISA definition of mathematical literacy but nevertheless are important prerequisites for it. In principle it is possible to possess mathematical literacy without harbouring such attitudes and emotions at the same time. In practice, however, it is not likely that mathematical literacy, as defined above, will be put into practice by someone who does not have self-confidence, curiosity, a feeling of interest, or the desire to do or understand things that contain mathematical components.

Organisation of the domain

For the purposes of OECD/PISA, it is useful to identify a number of aspects of mathematical literacy.

For OECD/PISA two major aspects and two minor aspects are used to organise the domain. The major aspects are:

– mathematical competencies; and

– mathematical big ideas.

The minor aspects are:

– mathematical curricular strands; and

– situations and contexts.

The major aspects are used for the purpose of describing the scope of the assessment and for describing proficiency. The minor aspects are used to ensure adequate coverage of the domain and balance in the range of assessment tasks selected.

It is important to point out that these four aspects should not be combined to form a single classification scheme. Two of the aspects, “mathematical big ideas” and “mathematical curricular strands”, are alternative schemes for describing mathematical content.

Mathematical competencies are general skills and competencies such as problem solving, the use of mathematical language and mathematical modelling.

Mathematical big ideas represent clusters of relevant, connected mathematical concepts that appear in *real* situations and contexts. Some of these big ideas are well established, such as *chance*, *change and growth*, *dependency and relationships* and *shape*. “Big ideas” are chosen because they do not result in the artificial approach of separating mathematics into different topics.

The aspect *mathematical curricular strands* represents the content of school mathematics as implemented in many school curricula. For OECD/PISA, *number, measurement, estimation, algebra, functions, geometry, probability, statistics,* and *discrete mathematics* are used. The mathematical curricular strands are identified as a minor aspect in OECD/PISA to ensure that the traditional curricular strands are adequately covered in the assessment. However, the actual selection of content for inclusion in OECD/PISA is driven by the more important and broader *mathematical big ideas* aspect.

The second minor aspect refers to *situations*, *i.e.* the settings in which the mathematical problems are presented. Examples are educational, occupational, public and personal settings.

The following provides a more detailed description of the four aspects.

Mathematical competencies

The first major aspect of the OECD/PISA mathematical literacy framework is mathematical competencies. This aspect is a non-hierarchical list of general mathematical skills that are relevant and pertinent to all levels of education. This list includes the following elements:

1. *Mathematical thinking skill*. This includes posing questions characteristic of mathematics ("Is there…?", "If so, how many?", "How do we find…?"); knowing the kinds of answers that mathematics offers to such questions; distinguishing between different kinds of statements (definitions, theorems, conjectures, hypotheses, examples, conditioned assertions); and understanding and handling the extent and limits of given mathematical concepts.
2. *Mathematical argumentation skill*. This includes knowing what mathematical proofs are and how they differ from other kinds of mathematical reasoning; following and assessing chains of mathematical arguments of different types; possessing a feel for heuristics ("What can(not) happen, and why?"); and creating mathematical arguments.
3. *Modelling skill*. This includes structuring the field or situation to be modelled; "mathematising" (translating "reality" into mathematical structures); "de-mathematising" (interpreting mathematical models in terms of "reality"); working with a mathematical model; validating the model; reflecting, analysing and offering a critique of a model and its results; communicating about the model and its results (including the limitations of such results); and monitoring and controlling the modelling process.
4. *Problem posing and solving skill*. This includes posing, formulating, and defining different kinds of mathematical problems ("pure", "applied", "open-ended" and "closed"); and solving different kinds of mathematical problems in a variety of ways.
5. *Representation skill*. This includes decoding, interpreting and distinguishing between different forms of representation of mathematical objects and situations and the interrelationships between the various representations; choosing, and switching between, different forms of representation, according to situation and purpose.
6. *Symbolic, formal and technical skill*. This includes: decoding and interpreting symbolic and formal language and understanding its relationship to natural language; translating from natural language to symbolic/formal language; handling statements and expressions containing symbols and formulae; using variables, solving equations and undertaking calculations.
7. *Communication skill*. This includes expressing oneself, in a variety of ways, on matters with a mathematical content, in oral as well as in written form, and understanding others' written or oral statements about such matters.
8. *Aids and tools skill*. This includes knowing about, and being able to make use of, various aids and tools (including information technology tools) that may assist mathematical activity, and knowing about the limitations of such aids and tools.

Competency classes

OECD/PISA does not propose the development of test items that assess the above skills individually. When doing *real* mathematics, it is usually necessary to draw simultaneously upon many (perhaps all) of the skills, so that any effort to assess individual skills is likely to result in artificial tasks and an unnecessary compartmentalisation of the mathematical literacy domain.

In order to operationalise this *mathematical competencies* aspect through the construction of items and tests, it is helpful to organise the skills into three larger classes of competency. The three competency classes are:

– Class 1: reproduction, definitions, and computations.

– Class 2: connections and integration for problem solving.

– Class 3: mathematical thinking, generalisation and insight.

Each of the skills listed above is likely to play a role in all competency classes. That is, the skills do not belong within only one competency class. The classes form a conceptual continuum, from simple reproduction of facts and computational skills, to the competency of making connections between different strands in order to solve simple real-world problems, and to the third class, which involves the "mathematisation" (this term is discussed in detail below) of real-world problems and reflection on the solutions in the context of the problems, using mathematical thinking, reasoning and generalisation.

The above discussion suggests that the classes form a hierarchy, in the sense that a set of tasks requiring Class 3 competencies will in general be more difficult than a set of tasks requiring Class 2 competencies. However, this does not imply that Class 2 competencies are a prerequisite for each Class 3 competency. In fact, previous studies (de Lange, 1987; Shafer and Romberg, in press) show that it is not necessary to excel in Class 1 competencies to do well in Class 2 or 3, while students performing well in Class 3 may not necessarily excel in Class 1 competencies.

The OECD/PISA definition of mathematical literacy places importance on students demonstrating the capacity to perform tasks requiring skills in all three competency classes. Assessment tasks will, therefore, be included from all three classes so that policy makers will have an opportunity to see how well their schools and curricula are developing the skills required in each of the competency classes.

Class 1 competencies: reproduction, definitions, computations

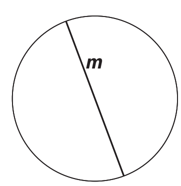
In this class, material typically featured in many standardised assessments and in comparative international studies is dealt with. The class includes knowledge of facts, representation, recognition of equivalents, recalling of mathematical objects and properties, performance of routine procedures, application of standard algorithms and development of technical skills. Manipulation of expressions containing symbols and formulae in standard form, and calculations, are also competencies in this class. Items that assess competencies in this class can usually be in a multiple-choice or a restricted open-ended format.

This class relates in particular to the *symbolic, formal and technical skill* that is described above. Some examples from this class are shown in Figure 2.

Figure 2. **Examples from Competency Class 1**

Solve the equation $7x - 3 = 13x + 15$
What is the average of 7, 12, 8, 14, 15, 9?
Write 69% as a fraction

Line ***m*** is called the circle's: __________



Class 2 competencies: connections and integration for problem solving

In this class, connections between the different strands and domains in mathematics are of importance, and information must be integrated in order to solve simple problems. Students will therefore have to choose which strategies and mathematical tools to use. Although the problems are classified as non-routine, they require only a relatively low level of mathematisation.

In this class, students are also expected to handle different methods of representation, according to situation and purpose. The connections component also requires students to be able to distinguish and relate different statements such as definitions, claims, examples, conditioned assertions and proofs.

The class relates to several of the mathematical skills mentioned above. It is clear that solving the problems given in the example requires some reasoning or argumentation; hence it requires the use of *mathematical argumentation skills*. Further, the students need to "model" the problem in order to solve it – thus *modelling skills* are required. The problem solving itself requires *problem-posing and problem-solving skills*. When in the process of solving the problem the students use various forms of representation – a table, a chart or a drawing – this requires *representation skills*.

From the mathematical language point of view, decoding and interpreting symbolic and formal language and understanding its relationship to natural language is another important skill in this class. Items in this class are often placed within a context, and engage students in mathematical decision making.

Two example problems from this class are shown in Figure 3. Unlike the examples from Class 1, it is not immediately clear to which curricular strand these questions belong, nor is it clear which method, strategy or algorithm would be best used to solve the problem. In fact, in some cases the curricular strand will depend upon the strategy which the student selects, and many alternative strategies may be equally suitable.

Figure 3. **Examples from Competency Class 2**

You have driven two thirds of the distance in your car. You started with a full fuel tank and your tank is now one quarter full. Do you have a problem?

Mary lives two kilometres from school, Martin five. How far do Mary and Martin live from each other?

Class 3 competencies: mathematical thinking, generalisation and insight

For items in this class, students are asked to "mathematise" situations, that is, to recognise and extract the mathematics embedded in the situation and to use mathematics to solve the problem; to analyse; to interpret; to develop their own models and strategies and to present mathematical arguments, including proofs and generalisations.

These competencies include an analysis of the model and reflection on the process. In this class of competencies, students should not only be able to solve problems but also to pose problems.

All these competencies will function well only if students are able to communicate adequately in different ways: oral, written, visual, etc. Communication is regarded as a two-way process: students should be able to communicate their mathematical ideas as well as to understand the mathematical communications of others.

Finally, it is important to stress that students also need *insight* into the nature of mathematics, including cultural and historical elements, and the use of mathematics in other contexts and other curriculum areas that are amenable to mathematical modelling.

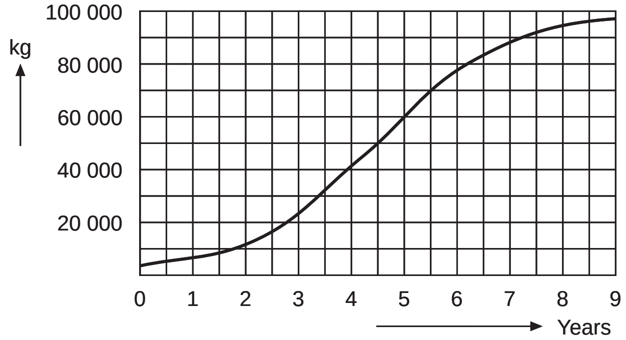
The competencies in this class often incorporate skills and competencies from other classes.

This class is a central component of mathematical literacy. Unfortunately, however, it is the most difficult class to assess, particularly in large-scale surveys such as OECD/PISA. Multiple-choice items, for example, are often not suitable for assessing these competencies. Extended-response questions with multiple answers are more likely to be an appropriate format, but both the design of such items and the assessment of students' responses to such items are very difficult. However, as this class forms a crucial part of mathematical literacy, as defined in OECD/PISA, an effort has been made to include it in the assessment, even though with only limited coverage.

An example problem from Class 3 is given in Figure 4.

Figure 4. **Examples from Competency Class 3**

Some fish were introduced to a waterway. The graph shows a model of the growth in the combined weight of fish in the waterway.



Suppose a fisherman plans to wait a number of years and then start catching fish from the waterway. How many years should the fisherman wait if he or she wishes to maximise the number of fish he or she can catch annually from that year on? Provide an argument to support your answer.

"Mathematisation"

Mathematisation, as it is used in OECD/PISA, refers to the organisation of perceived reality through the use of mathematical ideas and concepts. It is the organising activity according to which acquired knowledge and skills are used to discover unknown regularities, relationships and structures (Treffers and Goffree, 1985). This process is sometimes called *horizontal mathematisation* (Treffers, 1986). It requires activities such as:

- identifying the specific mathematics in a general context;
- schematising;
- formulating and visualising a problem;
- discovering relationships and regularities; and
- recognising similarities between different problems (de Lange, 1987).

As soon as the problem has been transformed into a mathematical problem, it can be resolved with mathematical tools. That is, mathematical tools can be applied to manipulate and refine the mathematically modelled real-world problem. This process is referred to as *vertical mathematisation* and can be recognised in the following activities:

- representing a relationship by means of a formula;
- proving regularities;
- refining and adjusting models;
- combining and integrating models; and
- generalising.

Thus the process of mathematisation occurs in two different phases: horizontal mathematisation, which is the process of translating the real world into the mathematical world, and vertical mathematisation, that is, working on a problem within the mathematical world and using mathematical tools in order to solve the problem. Reflecting on the solution with respect to the original problem is an essential step in the process of mathematisation that seldom receives adequate attention.

One can argue that mathematisation occurs in all competency classes because, in any contextualised problem, one needs to identify the relevant mathematics. However, in OECD/PISA, the kind of mathematisation that is required in Competency Class 3 is of particular importance. It is that form of mathematisation that goes beyond the mere recognition of well-known problems.

The differing complexity of mathematisation is reflected in the two examples below. Both are meant for students of 13 to 15 years of age and both draw upon similar mathematical concepts. The first requires simple mathematisation and the second requires more complex mathematisation.

The item in Figure 5 is an example of a Competency Class 2 item because it requires relatively low levels of mathematisation.

Figure 5. **Example item using simple "mathematisation"**

A class has 28 students. The ratio of girls to boys is 4:3. How many girls are in the class?

Source: TIMSS Mathematics Achievement in the Middle Years, p. 98.

The mathematisation required in the example in Figure 6 is from Competency Class 3 as it requires the student to recognise the relevant mathematics and to develop and then communicate a mathematical argument.

Figure 6. **Example item using more complex "mathematisation"**

In a certain country, the national defence budget is $30 million for 1980. The total budget for that year is $500 million. The following year the defence budget is $35 million, while the total budget is $605 million. Inflation during the period covered by the two budgets amounted to 10 per cent.

a) You are invited to give a lecture for a pacifist society. You intend to explain that the defence budget decreased over this period. Explain how you would do this.

b) You are invited to lecture to a military academy. You intend to explain that the defence budget increased over this period. Explain how you would do this.

Source: de Lange (1987); see also MSEB (1991). Used with permission.

Mathematical "big ideas"

As indicated earlier, the purposes of the OECD/PISA mathematics assessment are rather different from those of earlier comparative studies of mathematics, in particular IEA/TIMSS. Some of the most pertinent differences are described below.

The IEA/TIMSS test development process placed great emphasis on coverage of the curricula of participating countries and used a detailed scheme based upon traditional curriculum content strands to describe national curricula. However, school mathematics is often offered to students as a strictly compartmentalised science, and over-emphasises computation and formulae. Students leaving secondary school are typically not aware of the fact that mathematics is growing continuously, spreading into new fields and *situations*. As a result, the IEA/TIMSS instruments pertained mostly to knowledge of mathematical facts that were tested in isolation, mainly with very short items.

In contrast, OECD/PISA is focusing on mathematical literacy as it has been defined above. It is important, therefore, to emphasise that OECD/PISA's goal is to assess the full breadth of student achievement in a coherent, integrated way, rather than to test fragmented pieces of factual knowledge, which belongs to Competency Class 1. For OECD/PISA, interconnections and common ideas are central elements. Mathematics is the language that describes patterns, both patterns in nature and patterns invented by the human mind. In order to be mathematically literate, students must recognise these patterns and see their variety, regularity and interconnections.

It is for this reason that the traditional content strands are not a major dimension in the OECD/PISA mathematical literacy domain. Instead, the content to be assessed is organised around *big mathematical ideas.*

The concept of *"big ideas"* is not new. In 1990, the Mathematical Sciences Education Board (Senechal, 1990) published *On the Shoulders of the Giant: New Approaches to Numeracy*, which is a strong plea to help students delve deeper in order to find concepts that underlie mathematics and hence to reach a better understanding of their significance in the real world. For this it is necessary to explore ideas with deep roots in the mathematical sciences without concern for the limitations imposed by present school curricula. Other mathematicians support this idea, one of the better known publications being *Mathematics: The Science of Patterns* (Devlin, 1994, 1997).

A large number of big ideas can be identified. In fact, the domain of mathematics is so rich and varied that it would not be possible to draw up an exhaustive list of big ideas. For the purpose of focusing the OECD/PISA mathematical literacy domain, however, it is important that a selection of big ideas is made which encompasses sufficient variety and depth to reveal the essentials of mathematics.

The following list of mathematical big ideas is used in OECD/PISA to meet this requirement:

– chance;

– change and growth;

– space and shape;

– quantitative reasoning;

– uncertainty; and

– dependency and relationships.

In the first OECD/PISA assessment cycle, the limited testing time available for mathematics necessarily restricts the breadth of what can be assessed. The first cycle will therefore focus on the following two big ideas:

– change and growth; and

– space and shape.

There are two main reasons for limiting the first survey cycle to these two big ideas:

– first, these two domains cover a wide range of subjects from the content strands indicated earlier;

– second, these domains offer an adequate coverage of existing curricula.

Quantitative reasoning was omitted from the first survey cycle because of the concern that it would lead to an over-representation of typical number skills.

These two big ideas are elaborated further below.

Change and growth

Every natural phenomenon is a manifestation of change. Examples are: organisms changing as they grow, the cycle of seasons, the ebb and flow of tides, cycles of unemployment, weather changes and the Dow-Jones index. Some of these growth processes can be described or modelled by straightforward mathematical functions: linear, exponential, periodic, logistic, either discrete or continuous. But many processes fall into different categories and data analysis is quite often essential. The use of computer technology has resulted in more powerful approximation techniques and more sophisticated ways for visualising data. The patterns of change in nature and in mathematics do not follow the traditional content strands.

In order to be sensitive to the patterns of change, Stuart (1990) states that we need to:

– represent changes in a comprehensible form;

– understand the fundamental types of change;

– recognise particular types of change when they occur;

– apply these techniques to the outside world; and

– control a changing universe to our best advantage.

These competencies relate closely to both our definition of mathematical literacy and the competencies defined earlier in this framework.

Many different sub-strands of traditional content strands emerge in this big idea of change and growth. The obvious ones are relationships and functions, and their graphical representations. Series and gradients are closely intertwined with functions. An examination of rates of growth for different growth phenomena leads to linear, exponential, logarithmic, periodic and logistic growth curves, and to their properties and relationships. These, in turn, lead to aspects of number theory, such as Fibonacci numbers and the Golden Ratio. The connections between these ideas and geometrical representations can also play a role here.

Geometry can also be used to explore patterns in nature, art and architecture. Similarity and congruence might play a role, as would the growth of an area in relation to the growth of the perimeter or circumference.

Growth patterns can be expressed in algebraic forms, which in turn can be represented by graphs.

Growth can also be measured empirically, and questions arise as to what inferences can be drawn from the growth data, how the growth data may be represented, and so on. Aspects of data analysis and statistics are other content strands that come to light in this context.

Space and shape

Patterns are not only encountered in growth and change processes, but everywhere around us: spoken words, music, video, traffic, constructions and art. Shapes are patterns: houses, churches, bridges, starfish, snowflakes, town plans, clover leaves, crystals and shadows. Geometric patterns can serve as relatively simple models of many kinds of phenomena, and their study is possible and desirable at all levels (Grünbaum, 1985). Shape is a vital, growing, and fascinating theme in mathematics, with deep ties to traditional geometry (although relatively few to school geometry) but going far beyond it in content, meaning and method (Senechal, 1990).

In the study of shape and constructions, we are looking for similarities and differences as we analyse the components of form and recognise shapes in different representations and different dimensions. The study of shapes is closely connected to the concept of "grasping space" (Freudenthal, 1973). This means learning to know, explore and conquer, in order to live, breathe and move more easily in the space in which we live.

To achieve this, we must be able to understand the relative positions of objects. We must be aware of how we see things and why we see them as we do. We must learn to navigate through space and through constructions and shapes. This means that students should be able to understand the relationship between shapes and images or visual representations, such as that between a real city and photographs and maps of the same city. They must also understand how three-dimensional objects can be represented in two dimensions, how shadows are formed and must be interpreted, what perspective is and how it functions.

Described in this way, the study of space and shape is open-ended and dynamic, and fits well with mathematical literacy and the mathematical competencies defined for OECD/PISA.

Mathematical curricular strands

Of course one cannot, and should not, ignore traditional strands of the mathematics curriculum. This is why they are explicitly included as a minor organising aspect of the mathematical literacy domain in OECD/PISA. The *mathematical curricular strands* aspect can help to ensure a balance in the items and a reasonable spread of content in relation to the school curriculum. The content strands for OECD/PISA are:

– number;

– measurement;

– estimation;

– algebra;

– functions;

– geometry;

– probability;

– statistics; and

– discrete mathematics.

This list of curricular strands was developed in collaboration with countries participating in PISA. Items covering each of the above curricular strands will be included in the OECD/PISA assessments.

Situations

An important part of the definition of mathematical literacy is doing and using mathematics in a variety of situations. It has been recognised that the choice of mathematical methods and presentations of results is often dependent upon the settings in which the problems are presented. Each situation should allow students to participate in the process of mathematisation by recognising how practices learned in one situation can successfully be applied in other, similar situations.

One can think of *situations* as being at a certain "distance" from the students: of those identified for OECD/PISA, the closest is personal life, next is school life, work and sports (or leisure in general), followed by the local community and society as encountered in daily life, and furthest away are scientific contexts. Scientific contexts include proofs of abstract conjectures, generalisations of numeric or spatial patterns and the like.

In this way a more or less continuous scale has been identified that can be regarded as another aspect of the framework for OECD/PISA, in which the focus will be on five situations: personal, educational, occupational, public and scientific.

An aspect related to *situations* is the authenticity of the settings used in the problems. This aspect is discussed further below.

Task characteristics

In the previous sections, the OECD/PISA mathematical literacy domain has been defined and the structure of the domain has been described. This section considers the assessment tasks that will be used to assess the students. In this section the nature of the tasks, the mathematical context of the tasks, the task format, and the process of developing the tasks are described.

Contexts for items

An issue that needs to be considered when the assessment tasks are developed is the mathematical contexts in which the items are placed. The term *context* is used in accordance with the established use of this term in mathematics education. A context is an *extra*-mathematical or *intra*-mathematical setting within which the elements of a *mathematical complex* (*i.e.* a problem, a task or a collection of mathematical objects, relations, phenomena, etc.) are to be interpreted. A context is either a setting in which a given mathematical complex is already embedded (intra-mathematical setting), or a setting that lends itself to the activation of such a complex that then becomes embedded in that context (extra-mathematical setting). The embedding of a mathematical complex in an extra-mathematical context always implies the explicit or implicit (tacit) presence of a mathematical model which represents (aspects of) the setting by means of a translation into the mathematical complex at issue.

The above definition of context allows for a wide variety of contexts. For example, contexts can be drawn from other subjects, areas of professional or vocational practice, everyday life, life in the community and society, and so on. Leisure contexts such as sports and games are included in this definition as well. Situation, as defined earlier as a minor aspect, is one form of context. The pool of assessment tasks will use a variety of contexts. The variety is needed to ensure cultural diversity and to represent the range of roles that mathematics can play.

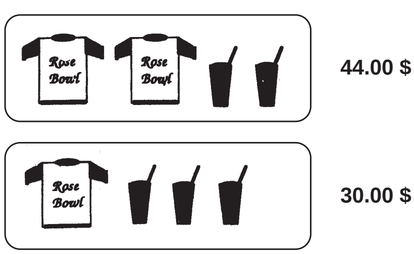
In OECD/PISA, the assessment tasks will focus on authentic contexts. A context is considered authentic if it resides in the actual experiences and practices of the participants in a real-world setting. Note that this definition does not require the students being assessed to be members of that setting. For example, questions about the yield from savings placed in a bank at a realistic interest rate may well be authentic even though they are outside the current sphere of experiences of the students who are being assessed.

It is important to recognise that using real components is not sufficient to make a context authentic. Consider, for example, the tasks in Figures 7 and 8. Both of these tasks include real elements but they are not authentic because no one in an out-of-school setting is likely to be called upon to address such problems.

The contexts of these problems have been chosen to make them look superficially like real-world problems. OECD/PISA avoids, wherever possible, such types of contexts.

The OECD/PISA emphasis on authentic contexts does not preclude the inclusion of important and/or interesting mathematical contexts (sometimes these may be virtual contexts). Consider, for example, the task in Figure 9. Here the context is stylised or generalised and it may or may not be authentic. Such

Figure 7. **A real but not authentic task**



How much does a T-shirt cost?
How much is a soda?
Show the reasoning that led you to these answers.

Figure 8. **Example item with a constructed context**

Which of the following numerical representations could be used to solve the problem? Bill weighed 107 pounds last summer. He lost 4 pounds, and then he gained 11 pounds. How much does he weigh now?

a) 107 - (4 + 11) = A
b) (107 - 4) + 11 = A
c) (107 + 11) + 4 = A
d) -4 + 11 = 107 + A
e) (107 - 11) + 4 = A

Figure 9. **Example item with a "virtual" context**

Would it be possible to establish a coinage system (or a stamp system) based on only the denominations 3 and 5? More specifically, what amounts could be reached on that basis? If possible, would such a system be desirable?

contexts will have a place in OECD/PISA if they are mathematically interesting and relevant. The use of mathematics to explain hypothetical scenarios and explore potential systems or situations, even if these cannot be carried out in reality, is one of its most powerful features.

Task formats

When assessment instruments are devised, the impact of the format of the tasks on student performance, and hence on the definition of the construct that is being assessed, must be carefully considered. This issue is particularly pertinent in a project such as OECD/PISA in which the large-scale cross-national context places serious constraints on the range of feasible item formats.

As in the case of the reading literacy domain, OECD/PISA will assess mathematical literacy through a combination of items with multiple-choice, closed-constructed response and open-constructed response formats. Appendix 2 discusses a broader range of formats that might be used when mathematics becomes a major domain in the second survey cycle.

Travers and Westbury (1989) state when discussing the second IEA mathematics study that: "The construction and selection of multiple-choice items was not difficult for the lower levels of cognitive behaviour – computation and comprehension." But, they continue, "difficulties were presented at the higher levels". There is a place for the use of the multiple-choice format (for an example see Figure 10), but only to a limited extent and only for the lowest goals (or behaviour) and learning outcomes. For any higher-order goals and more complex processes, other test formats should be preferred, the simplest being *open questions*.

Closed constructed-response items pose similar questions as multiple-choice items, but students are asked to produce a response that can be easily judged to be either correct or incorrect. When responses are not being machine-marked, this is a preferred format for assessing Competency Class 1, because guessing is not likely to be a concern, and the provision of distractors (which influence the construct that is being assessed) is not necessary. For example, for the problem in Figure 11 there is one correct answer and many possible incorrect answers.

Open constructed-response items require a more extended response from the student, and the process of producing a response is likely to involve higher-order activities. Often such items not only ask the student to produce a response, but also require the student to show the steps taken or to explain

Figure 10. **Example item with a limited number of response options**

A seal has to breathe even if it is asleep. Martin observed a seal for one hour. At the start of his observation the seal dove to the bottom of the sea and started to sleep. In 8 minutes it slowly floated to the surface and took a breath.

In 3 minutes it was back at the bottom of the sea again and the whole process started over in a very regular way.

After one hour the seal was:

a) at the bottom
b) on its way up
c) breathing
d) on its way down

Figure 11. **Example item with one correct answer and many incorrect answers**

Tepla Loroupe won the 1998 Rotterdam marathon. "It was easy", she said, "the course was quite flat".

Here you see a graph of the differences in elevation of the Rotterdam marathon course:

[***Differences in level of the course — in metres relative to the starting point***]

What was the difference between the highest and the lowest points of the course?

how the answer was reached. The key feature of open constructed-response items is that they allow students to demonstrate their abilities by providing solutions at a range of levels of mathematical complexity. The item in Figure 12 is an example.

For OECD/PISA, 25-35 per cent of the testing time devoted to mathematics will be allocated to open constructed-response items. These items require marking by a trained person who implements a marking rubric that may require an element of professional judgement. Because of the potential for disagreement between markers of these items, OECD/PISA will implement marker reliability studies to monitor the extent of disagreement. Experience in these types of studies shows that clear marking rubrics can be developed and reliable scores can be obtained.

The first survey cycle of OECD/PISA will make use of a task format in which several items are linked to common stimulus material. Tasks of this format give students the opportunity to become involved with a context or problem by asking a series of questions of increasing complexity. The first few questions are typically multiple-choice or closed-constructed items while subsequent items are typically open-constructed items. This format is appropriate for all competency classes.

Figure 12. **Example item which requires a constructed response**

Indonesia lies between Malaysia and Australia. Some data of the population of Indonesia and its distribution over the islands is shown in the following table:

Region	Surface area (km^2)	Percentage of total area	Population in 1980 (millions)	Percentage of total population
Java/Madura	132 187	6.95	91 281	61.87
Sumatra	473 606	24.86	27 981	18.99
Kalimantan (Borneo)	539 460	28.32	6 721	4.56
Sulawesi (Celebes)	189 216	9.93	10 377	7.04
Bali	5 561	0.30	2 470	1.68
Irian Jaya	421 981	22.16	1 145	5.02
TOTAL	1 905 569	100.00	147 384	100.00

One of the main challenges for Indonesia is the uneven distribution of the population over the islands. From the table we can see that Java, which has less than 7% of the total area, has almost 62% of the population.

Question: Design a graph (or graphs) that shows the uneven distribution of the Indonesian population.

Source: de Lange et Verhage (1992). Used with permission.

One reason for the use of such task formats is that it allows realistic tasks to be devised and the complexity of real-life situations to be reflected in them. Another reason relates to the efficient use of testing time, cutting down on the time required for a student to "get into" the subject matter of the situation. The necessity of making each scored point independent of others within the task is recognised and taken into account in the design of the OECD/PISA tasks. The importance to minimise bias that may be due to the use of fewer situations is also recognised.

Figure 13 is an example for such an item.

Assessment structure

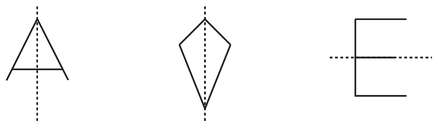
This section describes the structure of the mathematical literacy component of the OECD/PISA test booklets for the first assessment cycle, when a total of 60 minutes of testing time will be available for the assessment of mathematical literacy.

In the first survey cycle, testing time will be evenly distributed between the two mathematical big ideas: change and growth, and space and shape. The approximate division between the three competency classes will be 1:2:1. This information is summarised in Table 9, where the number of items for each big idea and competency class is given. The items are broken down according to whether they will require a single or multiple markers.

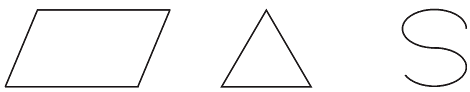
The minor aspects are not shown in Table 9. The distribution of the test items between the minor aspects, curricular strands and situations will be approximately uniform. That is, the OECD/PISA assessment will give equal emphasis to each of the nine curricular strands and the five situations mentioned earlier.

Figure 13. **A task with several items**

If a figure can be folded so that the two halves lie exactly on top of one another, the folding line is a line of symmetry.

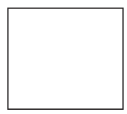


Question A
Which of the figures below have folding lines of symmetry?



(1) (2) (3)

Question B
Draw all the folding lines of symmetry on the square.



Question C
Which of the first eight capital letters of the alphabet have exactly two folding lines of symmetry?

Question D
John said: "I know a rule for being able to tell when a 4-sided figure has a folding line of symmetry. If the triangles on each side of the line have the same size and shape, it has a folding line of symmetry." Explain why you either agree or disagree with John.

Table 9. **Recommended number of items and score points across big ideas and competency classes**

Item type	Change and growth			Space and shape		
	Competencies			Competencies		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
One marker	6(6)	5(5)		6(6)	5(5)	
Multiple markers		2(5)	2(5)		2(5)	2(5)

Note: The numbers in parentheses are the expected number of score points.

Table 9 shows that the assessment is expected to contain:

– 15 items for each of the two big ideas;

– 21 score points for each of the two big ideas;

– 8 multiple-marker items and 22 single-marker items; and

– 12 score points in Competency Class 1, 20 score points in Competency Class 2, and 10 score points in Competency Class 3.

Over the longer term OECD/PISA will give greater emphasis to Competency Classes 2 and 3 than is the case in the first survey cycle, when testing time for mathematics will be very limited.

The following describes the composition of a typical half-hour module for the mathematical literacy assessment in the first cycle:

– a small number (2-4) of multiple-choice or closed-constructed response items assessing Competency Classes 1 or 2;

– a small number (1-2) of problems each containing two or three items in one context assessing Competency Classes 1 or 2;

– one *block* consisting of several items positioned in one context. The items would start with relatively straightforward tasks assessing Competency Class 1 and then progress to more complex ones assessing Competency Class 3.

Reporting scales

To meet the aims of OECD/PISA, the development of scales describing student achievement is essential. The process of developing the scales will be iterative, in that initial proposals that are based on past experience and research in the field of learning and cognitive development in mathematics will be further developed on the basis of empirical evidence collected during the PISA field trial.

The choice of reporting scales for mathematical literacy as a minor domain has yet to be made. The most obvious choices are to report: *i)* a single mathematical literacy scale; *ii)* a separate scale for each of the big ideas; or *iii)* a separate scale for each of the three competency classes. This choice between these three alternatives will be made after the field trial data has been analysed.

The results of OECD/PISA will be more informative if, for at least some of the items, marks are awarded not only for "correct" answers but also for the different strategies used by the students. It would not be surprising, for example, if students in two countries had the same score on the mathematical literacy scale, but differed greatly in how they had reached this level of mathematical literacy: in one country, students may rely more on formal strategies and routines, while in the other country more informal, common-sense strategies may be used.

Other issues

Links to other assessment areas

The focus of OECD/PISA is different from that of previous comparative surveys in mathematics, such as the Third International Mathematics and Science Study (IEA/TIMSS). While IEA/TIMSS is an instrument that was constructed around the common denominator of national curricula of participating countries, the aim of OECD/PISA is to assess mathematical literacy as defined above. Those familiar with IEA/TIMSS will see that there are relationships between OECD/PISA's "mathematical competencies" and IEA/TIMSS's "performance expectations", and that the "mathematical curricular strands" show certain similarities to those used in IEA/TIMSS. However, in OECD/PISA the curricular strands are subsumed as part of the big ideas that the domain of mathematical literacy addresses. Also, while most of the items in the IEA/TIMSS performance categories map on to Competency Class 1, OECD/PISA aims at covering Competency Classes 2 and 3 as well. For these reasons it is not expected that a psychometric link between the IEA/TIMSS and OECD/PISA reporting scales will be feasible.

Aids and tools

There are three possible policies with regard to the use of calculators and other tools that might be implemented in OECD/PISA:

– students could be prevented from using any calculators;

– students could be restricted to using a calculator supplied as part of the OECD/PISA assessment; or

– students could be free to use their own calculators and tools.

OECD/PISA will implement the third of the above possibilities. This option has been chosen because it represents the most authentic assessment of what students can achieve, and will provide the most informative comparison of the performance of education systems. A system's choice to allow students to access and use calculators is no different, in principle, from other instructional policy decisions that are made by systems and are not controlled by OECD/PISA.

Further, the argument that the first two options would lead to a fairer assessment, because of the apparent equality in the testing conditions, is superficial. Students who are used to answering questions with a calculator will be disadvantaged if this resource is taken away. Students who are provided with a calculator with which they are unfamiliar may not use that calculator efficiently, or its provision may encourage them to use it unnecessarily or inappropriately. For example, for many students, the simple problem $6 + 4 \times 3 = ?$ is more difficult with a simple calculator than without – particularly for those students who are not familiar with calculators.

It is therefore the policy of OECD/PISA that students should be allowed to use calculators and other tools as they are normally used in school. However, in OECD/PISA the test items will be chosen so that the use of calculators is not likely to enhance a student's performance in the assessment.

Chapter 3

SCIENTIFIC LITERACY

An important life skill for young people is the capacity to draw appropriate and guarded conclusions from evidence and information given to them, to criticise claims made by others on the basis of the evidence put forward, and to distinguish opinion from evidence-based statements. Science has a particular part to play here since it is concerned with rationality in testing ideas and theories against evidence from the world around. This is not to say that science excludes creativity and imagination, which have always played a central part in advancing human understanding of the world. Ideas which sometimes appear to have "come out of the blue" have been seized upon by a mechanism which Einstein described as "the way of intuition, which is helped by a feeling for the order lying behind the appearance" (Einstein, 1933). Which ideas are "seized upon" at a particular time has depended historically upon their social acceptability at that time, so that developments in scientific knowledge depend not only on the creativity of individuals but also on the culture in which they are proposed. But once the creative leap is made and a new theoretical framework for understanding has been articulated, then it has to be followed by painstaking testing against reality. As Hawking (1988) has written:

> "*A theory is a good theory if it satisfies two requirements: it must accurately describe a large class of observations on the basis of a model that contains only a few arbitrary elements, and it must make definite predictions about the results of future observations*" (Hawking, 1988, p. 9).

Theories that do not meet these requirements – or cannot be tested – are not scientific theories and it is important for an educated citizen to be able to distinguish between the kinds of questions that can be answered by science and those which cannot, and between what is scientific and what is pseudo-scientific.

Definition of the domain

Current thinking about the desired outcomes of science education for all citizens emphasises the development of a general understanding of important concepts and explanatory frameworks of science, of the methods by which science derives evidence to support claims for its knowledge, and of the strengths and limitations of science in the real world. It values the ability to apply this understanding to real situations involving science in which claims need to be assessed and decisions made. For example, Millar and Osborne (1998) have identified the focus of a modern science curriculum as being: "the ability to read and assimilate scientific and technical information and assess its significance". Their report continues:

> "*In this approach, the emphasis is not on how to 'do science'. It is not on how to create scientific knowledge, or to recall it briefly for a terminal examination. … Thus, in science, students should be asked to demonstrate a capacity to evaluate evidence; to distinguish theories from observations and to assess the level of certainty ascribed to the claims advanced*" (Millar and Osborne, 1998).

These should be the products of science education for all students. For some students, the minority who will become the scientists of tomorrow, this will be extended to in-depth study of scientific ideas and to the development of the ability to "do science".

With these points in mind, it is considered that the essential outcome of science education, which should be the focus OECD/PISA, is that students should be *scientifically literate*. This term has been used in different contexts. For example, the International Forum on Scientific and Technological Literacy for All (UNESCO, 1993) offered a variety of views, such as:

"*The capability to function with understanding and confidence, and at appropriate levels, in ways that bring about empowerment in the made world and in the world of scientific and technological ideas*" (UNESCO, 1993).

Included in the many different views of scientific literacy (reviewed by Shamos, 1995; see also Graeber and Bolte, 1997) are notions of levels of scientific literacy. For example, Bybee (1997) has proposed four levels, of which the lowest two are "nominal scientific literacy", consisting of knowledge of names and terms, and "functional literacy", which applies to those who can use scientific vocabulary in limited contexts. These are seen as being at levels too low to be aims within the OECD/PISA framework. The highest level identified by Bybee, "multidimensional scientific literacy", includes understanding of the nature of science and of its history and role in culture, at a level most appropriate for a scientific elite rather than for all citizens. It is, perhaps, the assumption that scientific literacy involves thinking at this level of specialisation that causes difficulty in communicating a more attainable notion of it. What is more appropriate for the purposes of the OECD/PISA science framework is closer to Bybee's third level, "conceptual and procedural scientific literacy".

Having considered a number of existing descriptions, OECD/PISA defines scientific literacy 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.*"

The following remarks further explain the meaning condensed in this statement.

Scientific literacy…

It is important to emphasise not only that both *scientific knowledge* (in the sense of knowledge about science) and the processes by which this knowledge is developed are essential for scientific literacy, but that they are bound together in this understanding of the term. As discussed in more detail below, the processes are only *scientific processes* when they are used in relation to the subject matter of science. Thus, using scientific processes necessarily involves some understanding of the scientific subject matter. The view of scientific literacy adopted here acknowledges this combination of ways of thinking about, and understanding, the scientific aspects of the world.

... use scientific knowledge to identify questions and to draw evidence-based conclusions…

In the above definition, *scientific knowledge* is used to mean far more than knowledge of facts, names and terms. It includes understanding of fundamental scientific concepts, the limitations of scientific knowledge and the nature of science as a human activity. The questions to be identified are those that can be answered by scientific enquiry, implying knowledge *about science* as well as about the scientific aspects of specific topics. Drawing *evidence-based conclusions* means knowing and applying processes of selecting and evaluating information and data, whilst 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.

... understand and help make decisions…

The phrase *understand and help make decisions* indicates first, that an understanding of the natural world is valued as a goal in itself as well as being necessary for decision-making and, second, that scientific understanding can contribute to, but rarely determines, decision-making. Practical decisions are always set in situations having social, political or economic dimensions and scientific knowledge is used in the context of human values related to these dimensions. Where there is agreement about the values in a situation, the use of scientific evidence can be non-controversial. Where values differ, the selection and use of scientific evidence in decision making will be more controversial.

... the natural world and the changes made to it through human activity

The phrase *the natural world* is used as shorthand for the physical setting, living things and the relationships among them. Decisions about the natural world include decisions associated with science related to self and family, community and global issues. *Changes made through human activity* refers to planned and unplanned adaptations of the natural world for human purposes (simple and complex technologies) and their consequences.

It is relevant to note here, and will be made more explicit later, that scientific literacy is not a dichotomy. That is, it is not suggested that people can be categorised as being either scientifically literate or scientifically illiterate. Rather, there is a progression from less developed to more developed scientific literacy. So, for example, the student with less developed scientific literacy might be able to identify some of the evidence that is relevant to evaluating a claim or supporting an argument or might be able to give a more complete evaluation in relation to simple and familiar situations. A more developed scientific literacy will show in more complete answers and the ability to use knowledge and to evaluate claims in relation to evidence in less familiar and more complex situations.

Organisation of the domain

The OECD/PISA definition of scientific literacy comprises three aspects:

- *scientific processes* which, because they are scientific, will involve knowledge of science, although in the assessment this knowledge must not form the major barrier to success;
- *scientific concepts*, the understanding of which will be assessed by application in certain content areas; and
- *situations* within which the assessment tasks are presented (this aspect is often referred to in common usage as the “context” or “setting”).

Although these aspects of scientific literacy are discussed separately it must be recognised that, in the assessment of scientific literacy, there will always be a combination of all three.

The first two of these aspects will be used both for the construction of tasks and for the characterisation of student performance. The third aspect will ensure that in the development of the assessment tasks due attention is paid to situating the science in a diverse range of relevant settings.

The following sections elaborate the three organising aspects. In laying out these aspects, the OECD/PISA framework has ensured that the focus of the assessment is upon the outcome of science education as a whole.

Scientific processes

Processes are mental (and sometimes physical) actions used in conceiving, obtaining, interpreting and using evidence or data to gain knowledge or understanding. Processes have to be used in relation to some subject matter; there is no meaning to a content-free process. They can be used in relation to a wide range of subject matter; they become *scientific processes* when the subject matter is drawn from scientific aspects of the world and the outcome of using them is to further scientific understanding.

What are commonly described as the processes of science range widely over the skills and understanding needed to collect and interpret evidence from the world around us and to draw conclusions from it. The processes relating to collecting evidence include those concerned with investigation in practice – planning and setting up experimental situations, taking measurements and making observations using appropriate instruments, etc. The development of these processes is included in the aims of school science education so that students can experience and understand the manner in which scientific understanding is built up and, ideally, the nature of scientific enquiry and of scientific knowledge. Few will require these practical skills in life after school but they will need the understanding of processes and concepts developed through practical, hands-on enquiry. Moreover, it has been strongly argued that what is traditionally regarded as the "scientific process", by which conclusions are drawn inductively from observations, and which is still reflected in much school science, is contrary to how scientific knowledge is developed (*e.g.* Ziman, 1980).

Scientific literacy, as identified here, gives higher priority to using scientific knowledge to "draw evidence-based conclusions" than to the ability to collect evidence for oneself. The ability to relate evidence or data to claims and conclusions is seen as central to what all citizens need in order to make judgements about the aspects of their life which are influenced by science. It follows that every citizen needs to know when scientific knowledge is relevant, distinguishing between questions which science can and cannot answer. Every citizen needs to be able to judge when evidence is valid, both in terms of its relevance and how it has been collected. Most important of all, however, every citizen needs to be able to relate evidence to conclusions based on it and to be able to weigh the evidence for and against particular courses of action that affect life at a personal, social or global level.

The distinctions that have just been made can be summarised briefly as giving priority to processes *about* science as compared with processes *within* science. It is important that the process skills listed in Figure 14 be read as being primarily about science and not primarily as they apply within science. All of the processes listed in Figure 14 involve knowledge of scientific concepts. In the first four processes this knowledge is necessary but not sufficient since knowledge about collecting and using scientific evidence and data is essential. In the fifth process the understanding of scientific concepts is the essential factor.

Figure 14. **Selected scientific processes**

1. Recognising scientifically investigable questions.
2. Identifying evidence needed in a scientific investigation.
3. Drawing or evaluating conclusions.
4. Communicating valid conclusions.
5. Demonstrating understanding of scientific concepts.

Some elaboration of these processes follows. They are further spelled out in operational terms later in Figure 19.

Recognising scientifically investigable questions

Recognising scientifically investigable questions can involve identifying the question or idea that was being (or could have been) tested in a given investigation. It may also involve distinguishing questions that can be answered by scientific investigation from those which cannot, or more openly suggesting a question that it would be possible to investigate scientifically in a given situation.

Identifying evidence needed in a scientific investigation

Identifying evidence needed in a scientific investigation involves identifying the information that is needed for a valid test of a given idea. This may require, for example, identifying or recognising 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.

Drawing or evaluating conclusions

Drawing conclusions or critically evaluating conclusions that have been drawn from given data may involve producing a conclusion from given scientific evidence or data or selecting from alternatives to the conclusion that fits the data. It may also involve giving reasons for or against a given conclusion in terms of the data provided or identifying the assumptions made in reaching a conclusion.

Communicating valid conclusions

Communicating to a specified audience valid conclusions from available evidence and data involves the production of an argument based on the situation and data given, or on relevant additional information, expressed in a manner that is appropriate and clear to the given audience.

Demonstrating understanding of scientific concepts

Demonstrating understanding of scientific concepts by applying appropriate concepts in a given situation involves explaining relationships and possible causes of given changes, or making predictions as to the effect of given changes, or identifying the factors that influence a given outcome, using scientific ideas and/or information which have not been given.

Some scientific knowledge is needed for all five processes. In the case of the first four, however, the knowledge is not intended to be the main "hurdle", since the aim is to assess the mental processes involved in gathering, evaluating and communicating valid scientific evidence. In the fifth process, on the other hand, it is the understanding of the scientific concept involved that is being assessed and this understanding is the main hurdle.

It is important to point out that, for each of the processes listed above, there is a wide range of task difficulties, depending upon the scientific concepts and situations involved. The OECD/PISA assessments will ensure that, through country feedback and the field trial, the items selected for the main study will be at the appropriate level of difficulty for 15-year-olds.

Concept and content

Concepts enable us to make sense of new experiences by linking them to what we already know. *Scientific concepts* are those which help to make sense of aspects of the natural and made worlds. Scientific concepts are expressed at many different levels, from the very broad labels of biological, physical, earth science, etc., under which they are presented in schools, to the long lists of generalisations such as often appear in statements of standards or curricula.

There are many ways of grouping scientific concepts to help the understanding of the scientific aspects of the world around. Sometimes concepts are labels indicating the range of characteristics that define a particular group of objects or events ("mammals", "acceleration", "solvent"); of these there must be several thousands. Concepts can also be expressed as generalisations about particular phenomena (the "laws" or theorems of physics or chemistry), of which there are many hundreds. They can also be expressed as major scientific themes which are more widely applicable and easier to operationalise for assessment and reporting purposes.

OECD/PISA uses four criteria for determining the selection of scientific concepts to be assessed:

– The first of these is relevance to everyday situations. Scientific concepts differ in the degree to which they are useful in everyday life. For example, although the theory of relativity gives a more accurate description of the relationships between length, mass, time and velocity, Newton's laws are more helpful in matters relating to the understanding of forces and motion encountered every day.

– The second criterion is that the concepts and content selected should have enduring relevance to life throughout the next decade and beyond. Given that the major assessment of science is planned to take place in the year 2006, the first cycle of OECD/PISA will focus on those concepts likely to remain important in science and public policy for a number of years.

– The third basis for selection is relevance to the situations identified as being ones in which scientific literacy should be demonstrated.

– The fourth criterion is that the concepts should require to be combined with selected scientific processes. This would not be the case where only recall of a label or of a definition was involved.

Figures 15 and 16 show the outcome of applying these criteria to the concepts and content of science. Figure 15 lists major scientific themes, with a few examples of the concepts relating to them. These broad concepts are what is required for understanding the natural world and for making sense of new experience. They depend upon and derive from study of specific phenomena and events but they go beyond the detailed knowledge that comes from study of these things. The concepts listed in Figure 15 are given to exemplify the meanings of the themes; there is no attempt to list comprehensively all the concepts which could be related to each theme.

The concepts given as examples in Figure 15 indicate that the knowledge that will be assessed relates to the major fields of science: physics, chemistry, biological sciences and earth and space sciences. Test items are classified by the major field of science as well as by the theme, area of application and process which they assess.

Figure 16 lists those areas of application of science that raise issues that the citizens of today and tomorrow need to understand and to make decisions about. It is these applications which guide the selection of content for tasks and items within them. Figure 16, therefore, indicates the areas of application in which the understanding of the concepts in Figure 15 will be assessed.

Figure 15. **Major scientific themes (with examples of related concepts) for the assessment of scientific literacy**

Structure and properties of matter
(thermal and electrical conductivity)

Atmospheric change
(radiation, transmission, pressure)

Chemical and physical changes
(states of matter, rates of reaction, decomposition)

Energy transformations
(energy conservation, energy degradation, photosynthesis)

Forces and movement
(balanced/unbalanced forces, velocity, acceleration, momentum)

Form and function
(cell, skeleton, adaptation)

Human biology
(health, hygiene, nutrition)

Physiological change
(hormones, electrolysis, neurons)

Biodiversity
(species, gene pool, evolution)

Genetic control
(dominance, inheritance)

Ecosystems
(food chains, sustainability)

The Earth and its place in the universe
(solar system, diurnal and seasonal changes)

Geological change
(continental drift, weathering)

Figure 16. **Areas of application of science for the assessment of scientific literacy**

Science in life and health Health, disease and nutrition Maintenance of and sustainable use of species Interdependence of physical/biological systems
Science in Earth and environment Pollution Production and loss of soil Weather and climate
Science in technology Biotechnology Use of materials and waste disposal Use of energy Transportation

As indicated earlier, OECD/PISA will include important concepts that are relevant to the science curricula of participating countries without being constrained by the common denominator of national curricula. In accordance with its focus on scientific literacy, it will do this by requiring application of selected scientific concepts and the use of scientific processes in important situations reflecting the real world and involving ideas of science.

Situations

Besides the processes and concepts assessed, the third feature of assessment tasks which affects performance is the situation in which the issues are presented. This is often called the *context* or *setting* of the tasks, but here the word *situation* is used to avoid confusion with other uses of these words. The particular situations are known to influence performance, so that it is important to decide and control the range of situations intended for the assessment tasks. It is not intended to report performance in relation to particular situations but they need to be identified in order to ensure a spread of tasks across those felt to be important and so that they can be controlled, as found necessary from field trials, from one survey to the next to ensure international comparability.

In selecting situations, it is important to keep in mind that the purpose of the assessment in science is to assess the ability of students to apply the skills and knowledge that they have acquired by the end of the compulsory years of schooling. OECD/PISA requires that the tasks should be framed in situations of life in general and not limited to life in school. In the school situation, scientific processes and concepts may be confined to the laboratory or classroom, but increasingly an attempt is being made also in countries' science curricula to apply these to the world outside the school.

Real-world situations involve problems which can affect us as individuals (*e.g.* food and energy use) or as members of a local community (*e.g.* treatment of the water supply or siting of a power station) or as world citizens (*e.g.* global warming, diminution of biodiversity). All of these are represented in the range of assessment tasks used in OECD/PISA. 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. In the framework of OECD/PISA the focus of the items will be on matters relating to the self and family (personal), to the community (public), to life across the world (global), and on those which illustrate how scientific knowledge evolves and affects social decisions associated with science (historical relevance).

In an international study it is important that the situations used for assessment items should be chosen in the light of relevance to students' interests and lives in all countries. They should also be appropriate for assessing scientific processes and concepts. Sensitivity to cultural differences has a high priority in task development and selection, not only for the sake of the validity of the assessment, but to respect the different values and traditions in participating countries. Feedback from field trials will be used to ensure that situations chosen for the survey tasks are relevant and appropriate across the different countries, whilst involving the combination of scientific knowledge with the use of scientific processes.

By setting test items in these situations OECD/PISA is seeking to assess the application of knowledge most likely to have been gained in the science curriculum (although some may be gained from other subjects and from non-school sources). However, although the knowledge that is required is curricular knowledge, in order to find out if this has gone beyond knowledge of isolated facts and is serving the development of scientific literacy, OECD/PISA is assessing the application of that knowledge in items reflecting real-life situations. Some of the examples of items presented below help to convey this point.

Task characteristics

In accordance with the OECD/PISA definition of scientific literacy, each assessment task will require the use of one or more of the processes in Figure 14 and, as has also been noted, some scientific knowledge. Tasks are envisaged as a series of questions (items) about stimulus material which presents the situation. Some tasks may include items assessing reading and mathematics as well as items relating to scientific literacy.

Examples of test items assessing scientific processes

Examples of the items being considered for assessing some of these processes will help to convey their operational meaning. The first two processes are assessed in two questions within a task entitled "Stop that germ!". The students are asked to read a short text which includes an extract about the history of immunisation. The extract on which two example questions are based is shown in Figure 17.

Figure 17. **Science Example 1**

As early as the 11th century, Chinese doctors were manipulating the immune system. By blowing pulverised scabs from a smallpox victim into their patients' nostrils, they could often induce a mild case of the disease that prevented a more severe onslaught later on. In the 1700s, people rubbed their skins with dried scabs to protect themselves from the disease. These primitive practices were introduced into England and the American colonies. In 1771 and 1772, during a smallpox epidemic, a Boston doctor named Zabdiel Boylston scratched the skin on his six-years-old son and 285 other people and rubbed pus from smallpox scabs into the wounds. All but six of his patients survived.

Example item 1: What idea might Zabdiel Boylston have been testing?

Example item 2: Give two other pieces of information that you would need to decide how successful Boylston's approach was.

Example Item 1 requires a constructed response which is scored as 2, 1 or 0 according to the amount of relevant detail given in the answer. (A score of 2 would be given to an idea along the lines of "breaking the skin and applying pus directly into the blood stream will increase the chances of developing immunity against smallpox".) The item assesses Process 1 – *Recognising scientifically investigable questions* using knowledge of *human biology* applied in the area of *science in life and health*.

Example Item 2 is also scored as 2, 1 or 0 according to whether one or both pieces of information are mentioned (the rate of survival without Boylston's treatment and whether his patients were exposed to smallpox apart from within the treatment). It assesses the Process 2 – *Identifying evidence needed in a scientific investigation* using knowledge of *human biology* and applied in the area of *science in life and health*.

The following four items are part of a task for which the stimulus material is a passage about Peter Cairney, who works for the Australian Road Research Board (Figure 18).

Example Item 3 assesses Process 2 – *Identifying evidence needed in a scientific investigation*, using knowledge of *forces and movement* in the area of *science in technology*.

Example Item 4 assesses Process 3 – *Drawing or evaluating conclusions*, using knowledge of *forces and movement* in the area of *science in technology*. No credit is given for agreeing or disagreeing but for the reason which is consistent with either and the given information (*e.g.* agree because there is less chance of

Figure 18. **Science Example 2**

... Another way that Peter gathers information is by the use of a TV camera on a 13 metre pole to film the traffic on a narrow road. The pictures tell the researchers such things as how fast the traffic is going, how far apart the cars travel, and what part of the road the traffic uses. Then after a time lane lines are painted on the road. The researchers can then use the TV camera to see whether the traffic is now different. Does the traffic now go faster or slower? Are the cars closer together or further apart than before? Do the motorists drive closer to the edge of the road or closer to the centre now that the lines are there? When Peter knows these things he can give advice about whether or not to paint lines on narrow roads.

Example item 3: If Peter wants to be sure that he is giving good advice, he might collect some other information as well beyond filming the narrow road. Which of these things would help him to be more sure about his advice concerning the effect of painting lines on narrow roads?

a) Doing the same on other narrow roads	Yes/No
b) Doing the same on wide roads	Yes/No
c) Checking the accident rates before and after painting the lines	Yes/No
d) Checking the number of cars using the road before and after painting the lines.	Yes/No

Scoring: Yes to *a)* and *c)* No to *b)* and *d)* (score 2)
Yes to *a)* No to *b)*, *c)* and *d)* (score 1)
Any other combination (score 0)

Example item 4: Suppose that on one stretch of narrow road Peter finds that after the lane lines are painted the traffic changes as in this table.

Speed	Traffic moves more quickly
Position	Traffic keeps nearer edges of road
Distance apart	No change

On the basis of these results it was decided that lane lines should be painted on all narrow roads.

Do you think this was the best decision?

Give your reasons for agreeing or disagreeing.

Agree

Disagree

Reason:

Example item 5: Drivers are advised to leave more space between their vehicle and the one in front when they are travelling more quickly than when they are travelling more slowly because faster cars take longer to stop.

Explain why a faster car takes longer to stop than a slower one.

Reason:

Example item 6: Watching his TV, Peter sees one car (A) travelling at 45 km/h being overtaken by another car(B) travelling at 60 km/h. How fast does car B appear to be travelling to someone in car A?

a) 0 km/h
b) 15 km/h
c) 45 km/h
d) 60 km/h
e) 105 km/h

collisions if the traffic is keeping near the edges of the road even if it is moving faster; if it is moving faster there is less incentive to overtake. Or, disagree because if the traffic is moving faster and keeping the same distance apart this may mean that they don't have enough room to stop in an emergency).

Example Item 5 assesses Process 5 – *Understanding of scientific concepts* about *forces and movement* in the area *of science in technology*. It requires a constructed response, marked 2, 1 or 0 according to whether one or both of the significant points are mentioned [reference to: *a)* greater momentum of a vehicle when it is moving more quickly and the consequent need for more force to stop it; *b)* at a higher speed a vehicle will move further whilst slowing down than a slower vehicle in the same time].

Example Item 6 assesses Process 5 – *Understanding of scientific concepts* about *forces and movement* in the area *of science in technology*. It is a straightforward multiple-choice item with only one correct answer *b)*, which is given one mark.

To answer all of these questions the student is required to use knowledge that would be gained from the science curriculum and apply it in a novel situation. Where assessment of conceptual understanding is not the main purpose of the item the knowledge required is not the main challenge (or hurdle) and success should depend on ability in the particular process required. Where assessment of conceptual understanding is the main aim, as in Example Items 5 and 6, the process is one of demonstrating this understanding.

Assessment structure

As the examples illustrate, what is identified as a defined "task" will take the form of several items linked to some initial stimulus material. Between them the items within a task may assess more than one process and one scientific concept, whilst each item assesses one of the scientific processes listed in Figure 14.

One reason for this structure is to make the tasks as realistic as possible and to reflect in them to some extent the complexity of real-life situations. Another reason relates to the efficient use of testing time, cutting down on the time required for a student to "get into" the subject matter of the situation, by having fewer situations, about which several questions can be posed rather than separate questions about a larger number of different situations. The necessity to make each scored point independent of others within the task is recognised and taken into account. It is also recognised that it is all the more important to minimise bias which may be due to the situation when fewer situations are used.

The tasks will be extended ones incorporating up to about eight items, each independently scored. In the great majority of tasks, if not all, there will be both items eliciting knowledge and understanding of the concepts involved, as in Example Items 5 and 6, and items requiring use of one or more of the processes of collecting and using evidence and data in a scientific investigation, as in Example Items 1-4. As indicated earlier, OECD/PISA will not include practical ("hands on") tasks, at least in the years 2000 and 2003, when science is a "minor" domain.

For the overall assessment, the desired balance between the processes is given in terms of percentages of scored points in Table 10. This may be revised for the assessment in 2006, when science will be the major domain of OECD/PISA.

Table 10. **Recommended distribution of score points across science processes**

Scientific processes	% scored points
Recognising scientifically investigable questions	10-15
Identifying evidence needed	15-20
Drawing or evaluating conclusions	15-20
Communicating valid conclusions	10-15
Demonstrating understanding of science concepts	40-50

It may well be that the topics of some tasks mean that the balance is tipped more towards assessment of understanding (Process 5), with the opposite occurring within other tasks. Where possible, items assessing Processes 1-4 and items assessing Process 5 will occur within each task, both to achieve the aim of covering important scientific concepts that students are likely to have developed from their school science curricula or outside school, and because the ability to use processes is very highly dependent upon the situation in which they are used (hence the processes need to be assessed in relation to a range of concepts). The aims of OECD/PISA suggest that both conceptual understanding and the combination of scientific knowledge with the ability to draw evidence-based conclusions are valued learning outcomes. The recommended target of roughly equal numbers of score points assigned to these two main kinds of learning outcomes should serve these aims.

As already noted, all types of items will be concerned with the application of scientific concepts that are likely to be developed in students through their school science curricula. Where the OECD/PISA science items differ from some – but by no means all – school science assessment is in their requirement that the concepts be applied in real-life situations. Similarly, the ability to draw evidence-based conclusions appears among the aims of many school science curricula. The OECD/PISA assessment will require the application of the processes in situations which go beyond the school laboratory or classroom. The extent to which this is novel to students will depend on how far applications in the real world are part of the curriculum they have experienced.

In relation to the areas of application, Table 11 shows that there will be as even a spread as possible across the three main groups.

Table 11. **Recommended distribution of score points across areas of application**

Areas of application of science	% scored points
Science in life and health	30-40
Science in earth and environment	30-40
Science in technology	30-40

In relation to the selection of situations, OECD/PISA will aim to spread the items evenly across the four identified situations: the personal, the community, the global and the historical.

The situations represented in the tasks will be defined by means of stimulus material, which may be a short written passage, or writing accompanying a table, chart, graph or diagram. The items will be a set of independently scored questions requiring the types of response indicated in Figure 19. Note that these include the ability to show understanding of scientific concepts by applying them in the areas of application. The responses required will relate to the situations and areas of application presented in the stimulus.

Until trials have taken place and the results have been analysed, it is not possible to be certain about, for example, the extent of uniformity across tasks in terms of the number of items within them, whether they relate to more than one area of application, or how items of different formats will be arranged. However we can summarise what is envisaged at this stage:

– All tasks will be extended, not single item tasks; they will include items assessing one or more of the scientific processes (Figure 14), knowledge of scientific concepts (Figure 15) and knowledge relating to one or more of the areas of application of science (Figure 16), and will require answers on paper (writing or drawing).

– Most tasks will be presented in written form, certainly for the survey in 2000, although the use of stimuli in other forms will be investigated for the year 2006, when science is the major element.

Figure 19. **Types of items for assessing scientific literacy**

Given	Response required
Recognising scientifically investigable questions	
An account of an investigation or procedure in which data were collected or comparisons were made	Select or produce the question or idea that was being (or could have been) tested
A description of a situation in which questions could be investigated scientifically	Formulate a question that could be investigated scientifically
Several questions/ideas arising from or relevant to the situation presented	Select the one(s) which can be answered/tested by scientific investigation
Identifying evidence needed in a scientific investigation	
An idea or hypothesis put forward in the question or stimulus material which is to be tested	Select or produce information about what is needed to test the idea or to make a prediction based on it. The information may be about: *a)* What things should be compared *b)* What variables should be changed or controlled *c)* What additional information is needed *d)* What action should be taken so that relevant data can be collected
Drawing or evaluating conclusions	
Data (test results, observations) from which conclusions can be drawn	Produce conclusion that fits the data
Data (test results, observations) and conclusions drawn from them	Select the conclusion that fits the data and give an explanation
Data (test results, observations) and a conclusion drawn from it	Produce reasons for the given data supporting or not supporting the conclusion or suggest the extent to which confidence can be placed in it
Communicating valid conclusions	
A situation in which (different) conclusions can be drawn or which requires information to be brought together to support a conclusion or recommendation and a specified audience	Produce an argument which is expressed clearly for the given audience and which is supported by relevant evidence/data found in the stimulus material
Demonstrating understanding of scientific concepts	
A situation in which a prediction, explanation or information is requested	Produce or select a prediction or explanation or additional information based on the understanding of a scientific concept or on information not given in the question or stimulus material

– Some tasks will include items which involve reading and/or mathematics and will contribute to the assessment in these domains. However there will be no items relating to scientific literacy that require only the repetition of information in the stimulus material, nor items that require only the recall of isolated factual information.

To cover the range of skills and understanding identified in this framework requires a range of item response formats. For example, multiple-choice items can be produced that validly assess those processes involving recognition or selection. However, for assessing the ability to evaluate and communicate, an open-response format is more likely to provide validity and authenticity. In many cases, however, the most appropriate format will depend on the particular content of the item.

Reporting scales

Scales and sub-scales

To meet the aims of OECD/PISA, the development of scales of student achievement is essential. The process of arriving at a scale has to be iterative, so that initial proposals, based on past experience of assessing science achievement and findings from research into learning and cognitive development in science, are modified by empirical findings from the OECD/PISA field trials.

Existing research and past experience suggest that there will be a scale of scientific literacy indicating development:

from being able to use scientific concepts that are more easily grasped and being able to do things such as the following, in familiar situations:

– recognise questions that can and those that cannot be decided by scientific investigation;

– identify information that has to be obtained in order to test a claim or explore an issue in situations where there is one variable to change and one to control;

– state why conclusions or claims may not be tenable in situations where there is no control of a variable that should have been controlled;

– present some of the main points in relating evidence to conclusions in a way that can be understood by others;

– make predictions and suggest explanations in terms of more easily grasped concepts;

... to being able to apply concepts of greater cognitive demand and do the following things, in more complex situations:

– recognise the tentativeness of all scientific understanding and the fact that testing of theories can lead to revision and better understanding;

– identify the information that has to be collected and the conditions under which it should be collected to test an explanation or explore an issue in complex situations;

– criticise the adequacy of information given in support of a claim or argument; argue for and against a statement or conclusion in relation to the evidence available in cases where there is no simple, clear causal relationship;

– present a well-constructed argument for and against a particular conclusion using scientific knowledge and data or information provided;

– make predictions and provide explanations based on understanding of more complex and abstract scientific concepts.

Details of the scientific literacy scale will emerge from analyses of results from field trials. These will show which items can be grouped together and which are spread at different points on the scale. The empirical data will be used to test the progression proposed here on the basis of judgement and what is already known about cognitive development.

In the year 2006, when the testing time available will enable a comprehensive coverage of scientific concepts and areas of application, it may be possible, in addition, to report a sub-scale of understanding of scientific concepts (Process 5), to be assessed by application in the situations presented. Such a scale will describe development *from* demonstrating correct but incomplete understanding, often involving concepts that are easier to grasp *to* demonstrating more complete understanding, often involving concepts of greater complexity.

In 2006 there will be sufficient information available across the scientific processes listed in Figure 14 to consider reporting sub-scales, which might relate, for example, to performance in separate processes or in the major fields of science. Again, this will depend on statistical, conceptual and policy considerations. If it proves feasible to report sub-scales, countries will have the benefit of being able to compare the achieved outcomes of their science education in detail with what they consider desirable outcomes.

Reporting on the content of, and incorrect responses to, different items is an important accompaniment to item statistics. It is expected that these content categories will be generated from the field trial and related to the kinds of answers actually given by students. Reporting some types of answers to specific items will also be necessary in order to illustrate the scale(s) and to give meaningful labels to it (them). This will involve releasing some items from those used in OECD/PISA.

Further levels of reporting are desirable and may become possible after the major science survey in 2006. One of these is performance in groups of items across tasks relating to the separate areas of application of science. This information will be useful in considering whether sufficient and effective attention is being given to issues of current concern.

Other issues

Links to other assessment areas

When information for a scientific literacy assessment task is presented in the form of an extended written passage, aspects of reading can be assessed. Similarly, when information is presented in the form of tables, charts, graphs, etc., the ability to read information can be assessed, and where some manipulation of number is required certain aspects of mathematics can be assessed. Such tasks will form part of the combined packages of the survey. Other tasks will assess only scientific processes involving drawing evidence-based conclusions and demonstrating understanding of scientific concepts.

A psychometric link between OECD/PISA and IEA/TIMSS will not be feasible for similar reasons to those explained in the chapter “Mathematical Literacy”.

Major and minor surveys in science

The surveys of science in the years 2000 and 2003, in which science is a “minor” element, will form the basis for comparisons over time. The restriction on the number of assessment tasks in 2000 and 2003 (even within a survey design which allows for different packages of items to be answered by different sub-samples of students) means that there will be fewer tasks relating to each area of application of science than will be possible in 2006. Thus the minor surveys of scientific literacy will involve the assessment of all the processes identified in Figure 14 and some of the concepts and areas of application identified in Figures 15 and 16. In the major year for science, 2006, a far more comprehensive coverage of the scientific concepts and areas of application will be possible.

REFERENCES

ALMOND, R.G. and MISLEVY, R.J. (1998),
Graphical Models and Computerized Adaptive Testing, TOEFL Technical Report No. 14, Educational Testing Service, Princeton, NJ, March.

BAKER, L. (1991),
"Metacognition, reading and science education", in C.M. Santa and D.E. Alvermann (eds.), *Science Learning: Processes and Applications,* International Reading Association, Newark, DE, pp. 2-13.

BENNETT, R.E. (1993),
"On the meanings of constructed response", in R.E. Bennett (ed.), *Construction vs. Choice in Cognitive Measurement: Issues in Constructed Response, Performance Testing, and Portfolio Assessment*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 1-27.

BINKLEY, M. and LINNAKYLÄ, P. (1997),
"Teaching reading in the United States and Finland", in M. Binkley, K. Rust and T. Williams (eds.), *Reading Literacy in an International Perspective,* US Department of Education, Washington, DC.

BRUNER, J. (1990),
Acts of Meaning, Harvard University Press, Cambridge, MA.

BYBEE, R.W. (1997),
"Towards an understanding of scientific literacy", in W. Grabe and C. Bolte (eds.), *Scientific Literacy – An International Symposium,* IPN, Kiel.

Council of Europe (1996),
Modern Languages: Learning, Teaching, Assessment. A Common European Framework of Reference, CC LANG (95) 5 Rev. IV, Strasbourg.

Council of Ministers of Education, Canada (1994),
Student Achievement Indicators Program: Reading and Writing, Toronto.

de LANGE, J. and VERHAGE, H. (1992),
Data Visualization, Sunburst, Pleasantville, NY.

de LANGE, J. (1987),
Mathematics, Insight and Meaning, OW and OC, Utrecht.

DEVLIN, K. (1994, 1997),
Mathematics, The Science of Patterns, Scientific American Library, New York.

DOLE, J., DUFFY, G., ROEHLER, L. and PEARSON, P. (1991),
"Moving from the old to the new: Research on reading comprehension instruction", *Review of Educational Research,* 16 (2), pp. 239-264.

EHRLICH, M.F. (1996),
"Metacognitive monitoring in the processing of anaphoric devices in skilled and less-skilled comprehenders", in C. Cornoldi and J. Oakhill (eds.), *Reading Comprehension Difficulties: Processes and Interventions,* Lawrence Erlbaum Associates, Mahwah, NJ, pp. 221-249.

EHRLICH, M.F., KURTZ-COSTES, B. and LORIDANT, C. (1993),
"Cognitive and motivational determinants of reading comprehension in good and poor readers", *Journal of Reading Behavior, 25,* pp. 365-381.

EINSTEIN, A. (1933),
"Preface to M. Plank", *Where is Science Going?,* Allen and Unwin, London.

ELLEY, W.B. (1992),
How in the World do Students Read?, International Association for the Evaluation of Educational Achievement, The Hague.

FREDERICKSON, N. (1984),
"The real test bias", *American Psychologist, 39,* pp. 193-202.

FREUDENTHAL, H. (1973),
Mathematics as an Educational Task, Reidel, Dordrecht.

FREUDENTHAL, H. (1983),
Didactical Phenomenology of Mathematical Structures, Reidel, Dordrecht.

GRAEBER, W. and BOLTE, C. (eds.) (1997),
Scientific Literacy – An International Symposium, IPN, Kiel.

GRONLUND, N.E. (1968),
Constructing Achievement Tests, Prentice Hall, Englewood Cliffs.

GRÜNBAUM, B. (1985),
"Geometry strikes again", *Mathematics Magazine,* 58 (1), pp. 12-18.

HAWKING, S.W. (1988),
A Brief History of Time, Bantam Press, London.

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

JONES, S. (1995),
"The practice(s) of literacy", in *Literacy, Economy and Society: Results of the First International Adult Literacy Survey,* OECD and Statistics Canada, Paris and Ottawa, pp. 87-113.

KIRSCH, I. (1995),
"Literacy performance on three scales: Definitions and results", in *Literacy, Economy and Society: Results of the First International Adult Literacy Survey,* OECD and Statistics Canada, Paris and Ottawa, pp. 27-53.

KIRSCH, I.S. and MOSENTHAL, P.B. (1989-1991),
"Understanding documents. A monthly column", *Journal of Reading,* International Reading Association, Newark, DE.

KIRSCH, I.S. and MOSENTHAL, P.B. (1994),
"Interpreting the IEA reading literacy scales", in M. Binkley, K. Rust and M. Winglee (eds.), *Methodological Issues in Comparative Educational Studies: The Case of the IEA Reading Literacy Study,* US Department of Education, National Center for Education Statistics, Washington, DC, pp. 135-192.

KIRSCH, I., JUNGEBLUT, A. and MOSENTHAL, P.B. (1998),
"The measurement of adult literacy", in T.S. Murray, I.S. Kirsch, and L. Jenkins (eds.), *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey,* US Department of Education, National Center for Education Statistics, Washington, DC.

LANGER, J. (1995),
Envisioning Literature, International Reading Association, Newark, DE.

LINNAKYLÄ, P. (1992),
"Recent trends in reading literacy research in Finland", in P. Belanger, C. Winter and A. Sutton (eds.), *Literacy and Basic Education in Europe on the Eve of the 21st Century,* Council of Europe, Strasbourg, pp. 129-135.

LUNDBERG, I. (1991),
"Reading as an individual and social skill", in I. Lundberg and T. Hoien (eds.), *Literacy in a World of Change,* Center for Reading Research/UNESCO, Stavanger.

MACCARTHEY, S.J. and RAPHAEL, T.E. (1989),
Alternative Perspectives of Reading/Writing Connections, College for Education, Institute for Research on Teaching. Occasional Paper #130, Michigan State University.

MILLAR, R. and OSBORNE, J. (1998),
Beyond 2000: Science Education for the Future, King's College London School of Education, London.

MYERS, M. and PARIS, S.G. (1978),
"Children's metacognitive knowledge about reading", *Journal of Educational Psychology, 70,* pp. 680-690.

PARIS, S., WASIK, B. and TURNER, J. (1991),
"The development of strategic readers", in R. Barr, M. Kamil and P. Mosenthal (eds.), *Handbook of Reading Research*, Vol. II, Longman, New York.

SENECHAL, M. (1990),
"Shape", in L.A. Steen (ed.), *On the Shoulders of the Giant – New Approaches to Numeracy*, National Academy Press, Washington, DC, pp. 139-182.

SHAFER, M.C. and ROMBERG, T. A (in press),
"Assessment in Classroom that Promote Understanding", in E. Fennema and T.A. Romberg (eds.), *Mathematics Classrooms that Promote Understanding,* Erlbaum, Mahwah, NJ.

SHAMOS, M.H. (1995),
The Myth of Scientific Literacy, Rutgers University Press, New Brunswick.

SMITH, M.C. (1996),
"Differences in adults' reading practices and literacy proficiencies", *Reading Research Quarterly, 31,* pp. 196-219.

STICHT, T.G. (ed.) (1975),
Reading for Working: A Functional Literacy Anthology, Human Resources Research Organization, Alexandria, VA.

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

STREEFLAND, L. (1990),
Fractions in Realistic Mathematics Education, A Paradigm of Developmental Research, Reidel Dordrecht.

STUART, I. (1990),
"Change", in L.A. Steen (ed.), *On the Shoulders of the Giant – New Approaches to Numeracy*, National Academy Press, Washington, DC, pp. 183-218.

TAUBE, K. and MEJDING, J. (1997),
"A nine-country study: What were the differences between the low and high performing students in the IEA Reading Literacy Study?", in M. Binkley, K. Rust and T. Williams (eds.), *Reading Literacy in the International Perspectives,* US Department of Education, National Center for Education Statistics, Washington, DC, pp. 63-100.

TRAUB, R.E. (1993),
"On the equivalence of the traits assessed by multiple-choice and constructed-response tests", in R.E. Bennett (ed.), *Construction vs. Choice in Cognitive Measurement: Issues in Constructed Response, Performance Testing, and Portfolio Assessment*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 29-44.

TRAVERS, K.J. and WESTBURY, I. (1989),
The IEA Study of Mathematics, Vol. 1, Analysis of mathematics curricula, Pergamon Press, Oxford.

TREFFERS, A. (1986),
Three Dimensions, Reidel, Dordrecht.

TREFFERS, A. and GOFFREE, F. (1985),
"Rational analysis of realistic mathematics education", in L. Streefland (ed.), *Proceedings of the Ninth International Conference for the Psychology of Mathematics Education (PME),* OW and OC, Utrecht, pp. 79-122.

UNESCO (1993),
International Forum on Scientific and Technological Literacy for All, Final Report, UNESCO, Paris.

WARD, W.C., DUPREE, D. and CARLSON, S.B. (1987),
A Comparison of Free-response and Multiple-choice Questions in the Assessment of Reading Comprehension (RR-87-20), Educational Testing Service, Princeton, NJ.

WERLICH, E. (1976),
A Text Grammar of English, Quelle and Meyer, Heidelberg.

ZIMAN, J.M. (1980),
Teaching and Learning about Science and Society, Cambridge University Press.

Appendix 1

FUNCTIONAL EXPERT GROUP MEMBERSHIP

Reading

Irwin Kirsch, Chair
Educational Testing Service
Princeton, New Jersey, United States

Marilyn Binkley
National Center for Education Statistics
Washington, DC, United States

Alan Davies
University of Edinburgh
Scotland, United Kingdom

Stan Jones
Statistics Canada
Nova Scotia, Canada

John de Jong
CITO, National Institute for Educational Measurement
Arnhem, Netherlands

Dominique Lafontaine
Université de Liège
Liège, Belgium

Pirjo Linnakylä
University of Jyvaskyla
Jyvaskyla, Finland

Martine Rémond
Institut National de Recherche Pédagogique
Paris, France

Wolfgang Schneider
University of Würzburg
Würzburg, Germany

Ryo Watanabe
National Institute for Educational Research
Tokyo, Japan

Mathematics

Jan de Lange, Chair
Utrecht University
Utrecht, Netherlands

Raimondo Bolletta
Centro Europeo dell'Educazione
Frascati, Italy

Sean Close
St. Patricks College
Dublin, Ireland

Maria Luisa Moreno
Instituto Nacional de Calidad y Evaluación (INCE)
Madrid, Spain

Mogens Niss
Roskilde University
Roskilde, Denmark

Kyung Mee Park
Korea Institute of Curriculum and Evaluation
Seoul, Korea

Thomas Romberg
University of Wisconsin-Madison
Madison, Wisconsin, United States

Peter Schüller
Federal Ministry of Education and Cultural Affairs
Vienna, Austria

Science

Wynne Harlen, Chair
The Scottish Council for Research in Education
Edinburgh, United Kingdom

Peter Fensham
Monash University
Melbourne, Australia

Raul Gagliardi
Geneva, Switzerland

Donghee Shin
Korea Institute of Curriculum and Evaluation
Seoul, Korea

Svein Lie
University of Oslo
Oslo, Norway

Manfred Prenzel
Institut für die Pädagogik der Naturwissenschaften (IPN)
(Institute for Science Education)
Kiel, Germany

Senta Raizen
National Center for Improving Science Education
Washington, DC, United States

Elizabeth Stage
University of California
Oakland, California, United States

Appendix 2

CONSIDERATIONS FOR FUTURE SURVEY CYCLES OF OECD/PISA

Exposure and attitudes relevant to the assessment of scientific literacy

In addition to the cognitive outcomes it is important to collect information about the affective outcomes of science education. Many different types of information concerning students' participation in science-based activities and their views about the value of scientific activity for themselves and for society in general are relevant to OECD/PISA. Furthermore, since the tasks used in OECD/PISA surveys may be novel to some students it will be important to have information about familiarity with the content and form of the tasks and items. In some cases this information can provide explanatory variables; in other cases it can significantly add to the data about the outcomes of students' science education.

In future cycles it is therefore planned to collect information, either in the student questionnaire or within a test package, about students':

– participation in science at school and outside school (*e.g.* their participation in extra-curricular science activities, such as reading science magazines, watching science television programmes and involvement in community activities such as environmental organisations);

– evaluation of the utility of the science they have learned in school and out of school for making personal and community decisions;

– judgements about the role of science in creating or solving problems;

– exposure to learning about the particular topics on which the assessment tasks have been based;

– familiarity with the form of the task and items.

This information should be gathered in the general student questionnaire, with the exception of the fourth item: exposure to learning about the particular topics on which the assessment tasks have been based. This could be posed as a final question in the science tests. The question would list the test items and ask students to indicate for each item whether they had heard some, little or no discussion on a topic related to the question, first, in school and, second, in a non-school setting.

Alternative item formats in mathematics

When mathematics becomes a major domain in OECD/PISA in 2003, it may be possible to extend the range of assessment tasks that can be used. This will be particularly important if future cycles intend to pay greater attention to the assessment of tasks in Competency Class 3.

Extended response essay tasks

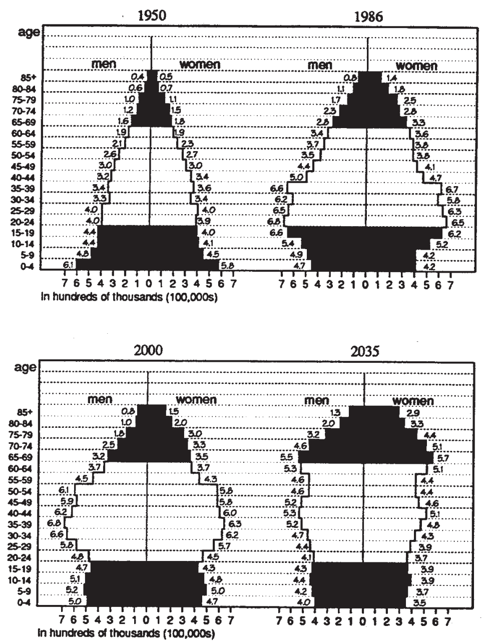
One possibility is to use extended-response essay tasks. According to Gronlund (1968), extended-response essay tasks are inefficient for measuring knowledge outcomes, but they provide a freedom of response which is needed for measuring complex outcomes. These complex outcomes include the ability to create, to organise, to integrate, to express, and similar types of action and behaviour that call for the production and synthesis of ideas.

Extended-response essay tasks give students an opportunity to probe more deeply into a complex problem. In contrast to a series of items linked to a common stimulus, such tasks allow students to develop their own responses as they examine a problem. This is a particularly appropriate format for assessing Competency Class 3.

Figure 20 provides an example of an extended-response essay task.

Figure 20. **A mathematics essay task**

The four pyramids show the population of the Netherlands.
The first two show the actual population in 1950 and 1986.
The last two show the expected population for 2000 and 2035.



Discuss the extent to which the information provided in these population pyramids suggests an aging of the population in the Netherlands; where appropriate use alternative graphical displays to support your argument.

Oral tasks

In some countries, oral assessment is or has been common practice, even as part of national examinations. There are different forms of oral assessment, such as:

– an oral discussion on certain mathematical subjects that are known to the students;

– an oral discussion on a subject, covering a task to be done at home, that is given to the students prior to the discussion; and

– an oral discussion on a task to be done at home after the assessment has been completed by the students.

The oral assessment format is frequently used to operationalise higher-order processes.

Two-stage tasks

A combination of different task formats can be referred to as a two-stage task. A written task followed by an oral task on the same subject is a typical example. Two-stage tasks combine the advantages of traditional written tests with a set time limit with the opportunities provided by open tasks.

The following describes a two-stage task. The first stage comprises a written assessment with the following characteristics:

– all students are given the assessment at the same time;

– all students must complete the assessment within a fixed time limit;
– the assessment is oriented more towards finding out what students do not know rather than what they do know;
– usually the assessment is based on the "lower" goal activities of computation and comprehension;
– the assessment consists of open questions; and
– marks are as objective as they can be.

The second stage compensates for the elements missing from the first stage. The characteristics of the second stage are that:

– the assessment is undertaken with no time limit;
– the assessment may be completed at home;
– the assessment emphasises what students know rather than what they do not know;
– much attention is given to higher-order activities: interpretation, reflection, communication, etc.;
– the structure of the assessment is more open, consisting of questions calling for long answers, including essays; and
– marking can be difficult and less objective.

Production items

Assessment procedures should provide students with the opportunity to demonstrate their abilities, and should be considered an integral part of the learning-teaching process. The use of students' *own productions* is not new, and there exists considerable experience with this form of assessment. Treffers (1987) introduced a distinction between construction and production. Free construction offers the greatest scope in terms of freedom for students to demonstrate their own thoughts and abilities. It can involve:

– solving relatively open problems which invite the production of divergent responses because they admit different solutions, often at varying levels of mathematisation;
– solving incomplete problems which, before they can be solved, require students to search for and obtain data and references;
– contriving one's own problems (easy, moderate, difficult), written in the form of a test paper or a book of problems about a topic for the next cohort of students (Streefland, 1990).

These possible formats, and others, should be explored further for use in OECD/PISA in the years to come.

OECD PUBLICATIONS, 2, rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(96 1999 05 1 P) ISBN 92-64-17053-7 – No. 50619 1999