



OECD Education Working Papers No. 102

Comparison of PIAAC  
and PISA Frameworks  
for Numeracy  
and Mathematical Literacy

**Iddo Gal,  
Dave Tout**

<https://dx.doi.org/10.1787/5jz3wl63cs6f-en>

**Unclassified**

**EDU/WKP(2014)1**

Organisation de Coopération et de Développement Économiques  
Organisation for Economic Co-operation and Development

**27-May-2014**

**English text only**

**DIRECTORATE FOR EDUCATION AND SKILLS**

**Cancels & replaces the same document of 22 May 2014**

**COMPARISON OF PIAAC AND PISA FRAMEWORKS FOR NUMERACY AND MATHEMATICAL LITERACY**

**OECD Education Working Paper n°102**

*This paper has been prepared by:*  
*- Iddo Gal (University of Haifa, Israel)*  
*- David Tout (Australian Council for Educational Research)*

**JT03358132**

**Complete document available on OLIS in its original format**

*This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.*



**EDU/WKP(2014)1**  
**Unclassified**

**English text only**

## OECD EDUCATION WORKING PAPERS SERIES

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed herein are those of the author(s).

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcome, and may be sent to the Directorate for Education and Skills, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgement of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to *rights@oecd.org*.

Comment on the series is welcome, and should be sent to *edu.contact@oecd.org*.

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

---

[www.oecd.org/edu/workingpapers](http://www.oecd.org/edu/workingpapers)

---

Copyright © OECD 2014.

## **COMPARISON OF PIAAC AND PISA FRAMEWORKS FOR NUMERACY AND MATHEMATICAL LITERACY**

Iddo Gal  
University of Haifa, Israel  
iddo@research.haifa.ac.il

Dave Tout  
Australian Council for Educational Research  
david.tout@acer.edu.au

Key words: Numeracy, Mathematical literacy, Large-scale assessment, Competencies

### **Abstract**

This paper describes key aspects of the frameworks for the assessment of adult numeracy and mathematical literacy in PIAAC and PISA, which are OECD two flagship programs for international comparative assessment of competencies. The paper examines commonalities and differences in how the constructs of adult numeracy and mathematical literacy were assessed in PIAAC and PISA, and sketches selected challenges associated with interpretation of results from these surveys.

### **Résumé**

Ce document décrit les principaux aspects des cadres d'évaluation de la numératie des adultes et de la culture mathématique dans PIAAC et PISA, deux programmes phares de l'OCDE pour l'évaluation comparative internationale des compétences. Le document examine les points communs et les différences dans la façon dont les concepts de numératie des adultes et de la culture mathématique ont été évalués dans PISA et PIAAC, et esquisse quelques défis associés à l'interprétation des résultats de ces enquêtes.

## TABLE OF CONTENTS

Preface.....	6
CHAPTER 1: A POLICY PERSPECTIVE ON NUMERACY AND MATHEMATICAL LITERACY ASSESSMENTS.....	8
CHAPTER 2: NUMERACY AND MATHEMATICAL LITERACY: CONCEPTUAL FRAMEWORKS.....	11
2.1 Mathematical literacy and PISA.....	12
2.2 Numeracy and PIAAC.....	14
2.3 Summary .....	17
CHAPTER 3: METHODOLOGIES IN PIAAC AND PISA AND THEIR IMPACT ON THE ASSESSMENT OF MATHEMATICAL LITERACY AND NUMERACY .....	18
3.1 Overall approach and sampling and time issues.....	18
3.2 Test design in PIAAC.....	19
3.3 Test characteristics in PISA .....	21
3.4 Scoring .....	22
3.5 Correlates and background questionnaires.....	22
3.6 Measurement scales and comparability of reported proficiencies.....	23
3.7 Summary .....	24
CHAPTER 4: COMPARISONS OF FRAMEWORKS AND COVERAGE .....	26
4.1 Mathematical content .....	26
4.1.1 PIAAC and PISA content classifications.....	26
4.1.2 Other content related issues .....	29
4.2 Contexts.....	30
4.3 Responses/actions.....	31
4.3.1 PIAAC response classifications.....	32
4.3.2 PISA response classifications 2003 to 2009 .....	32
4.3.3 PISA 2012 response classifications – the three mathematical processes.....	33
4.3.4 Summary and comparison of responses/actions .....	33
4.4 Item formats .....	34
4.4.1 PISA item formats .....	34
4.4.2 PIAAC item formats .....	36
4.4.3 Summary and comparison of item formats.....	36
4.5 Representations, reading demands, and authenticity-related issues .....	38
4.5.1 Representations.....	38
4.5.2 Reading demands.....	38
4.5.3 Authenticity of items and tasks.....	39
4.6 Item complexity factors.....	40
4.6.1 PIAAC complexity scheme .....	40
4.6.2 PISA Fundamental Mathematical Capabilities and item complexity factors .....	42
4.6.3 Summary and comparison of item complexity factors. ....	43
4.7 Sample items .....	43

CHAPTER 5: SUMMARY AND CONCLUSIONS .....	47
5.1 Conceptual issues: commonalities and differences .....	47
5.2 Assessment and methodology issues: commonalities and differences.....	50
5.2.1 The number, characteristics, and range of assessment items .....	50
5.2.2 Comparability of reported results .....	51
5.3 Conclusions .....	52
REFERENCES .....	53
RECENT OECD PUBLICATIONS OF RELEVANCE TO THIS WORKING PAPER.....	57
THE OECD EDUCATION WORKING PAPERS SERIES ON LINE.....	57

## COMPARISON OF PIAAC AND PISA FRAMEWORKS FOR NUMERACY AND MATHEMATICAL LITERACY

### Preface

OECD is presently conducting in parallel two major international comparative assessments of key competencies, the Program for International Student Assessment (PISA) which is focused on competencies of 15-year-old students in close to 70 countries, and the Programme for International Assessment of Adult Competencies (PIAAC), which focuses on competencies of adults in ages 16-65 in close to 35 countries. Both assessments are planned on a cyclic basis, every three years for PISA, every five to ten years for PIAAC. This paper examines the commonalities and differences of the Mathematical Literacy (ML) framework for PISA and the Numeracy framework for PIAAC, which assess related constructs of interest to diverse stakeholders in many countries.

“Numeracy” and “mathematical literacy” are constructs that in general terms pertain to the ability of individuals, whether school students or adults, to cope with tasks that are likely to appear in the adult world and that contain mathematical or quantitative information, or that require the activation of mathematical or statistical skills and knowledge. Both constructs (and related terms such as quantitative literacy, statistical literacy, and others) have received and continue to receive significant attention from policy makers and many other stakeholders, given the importance of adult numeracy and mathematical literacy for effective functioning of individuals in personal and community life and in labor market and further learning, and for the well-being of citizens, societies, and economies (European Commission, 1996; Steen, 2001; Hoyles, Wolf, Molyneux-Hodson, & Kent, 2002).

However, although numeracy and mathematical literacy are related constructs, they are not simply different names for the same underlying entity. Rather, they reflect ideas that have evolved along different trajectories, in the case of numeracy well outside large-scale assessment but in broad educational contexts. Further, when viewed in the context of large-scale assessment programs, *results* reported about the proficiency of individuals in numeracy and mathematical literacy cannot be automatically viewed as reflecting the same underlying entities (i.e. knowledge or skills), because such results are shaped by the different characteristics and constraints of the actual assessment methodologies. Yet, irrespective of methodological differences between assessments, decisions (e.g. by a government ministry, by a school system) about social policy or educational interventions *have to go back to the meaning of the underlying constructs*. Hence, a conceptual analysis of numeracy and mathematical literacy is essential since it can inform policy and interventions just as much as actual assessment results.

PISA frameworks and results have been released in multiple cycles since 2000 and PISA has gained much recognition worldwide, thus when policy makers and educators think of “mathematical literacy” they naturally think of PISA-related constructs. Several analyses of the PISA mathematical literacy framework and results have been published. For example, mathematics assessments in PISA have been compared to those in Trends in International Mathematics & Science Study (TIMSS) (e.g. Ferrini-Mundy & Schmidt, 2005; Ginsburg, Cooke, Leinwand, Noell & Pollock, 2005; Wu, 2010), or in the National Assessment of Educational Progress (NAEP, USA; Neidorf, Binkley, Gattis & Nohara, 2006). However, because mathematical literacy is the major domain in PISA 2012, the mathematical literacy framework in PISA 2012 has been revised compared to previous cycles of PISA.

PIAAC is still underway: results of its first wave, based on 24 countries, were released in October 2013 (OECD, 2013a), with 9 additional countries following in a second wave. While PIAAC may seem to many like a new effort, it builds on the solid foundations of over 30 years of national and

international surveys of adult skills, and has been designed on the basis of prior assessments and lessons learned from the Adult Literacy and Lifeskills survey (ALL), the International Adult Literacy Survey (IALS) (Statistics Canada and OECD, 1996; 2005), and earlier studies.

With the above in mind, this paper has been prepared to assist policy makers, educators, researchers, and other stakeholders, in understanding the commonalities and differences of the mathematical literacy framework for PISA (including from 2000 through to the latest PISA 2012 version) and the numeracy framework for PIAAC. Such a comparison may help in interpreting results from PISA and PIAAC, and could inform further research and development regarding education and training of the respective target populations of school-age learners and adults. This paper presents a comparison that is quite conceptual in nature as results from PISA 2012 had not been released at the time of writing.

With the above in mind, this paper is organised in five main chapters, as follows: Chapter 1 outlines the goals of PIAAC and PISA. Chapter 2 reviews the conceptualisation of mathematical literacy in PISA and numeracy in PIAAC. Chapter 3 presents further technical information about the PIAAC and PISA assessment programmes, and compares how the assessment of mathematical literacy and numeracy are implemented in these programs. Chapter 4 elaborates on details of the assessment scales and items used in each assessment. Finally, Chapter 5 summarises selected key points and presents some conclusions emerging from the analysis.



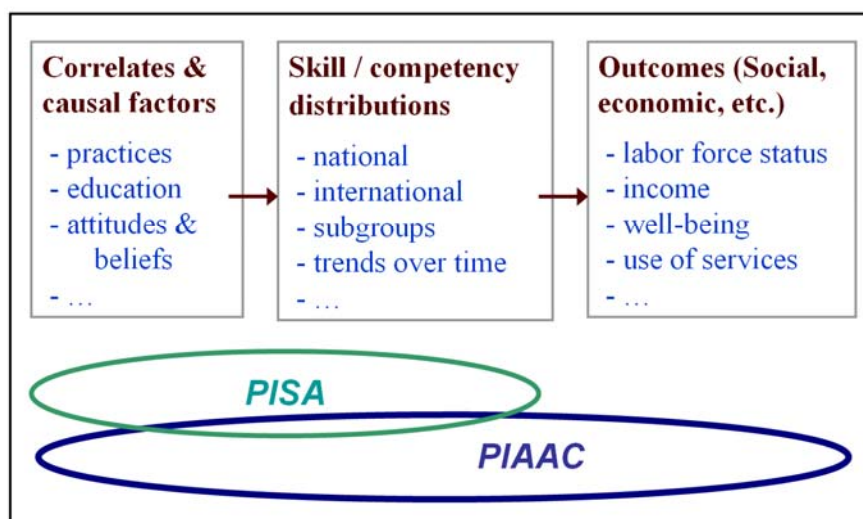
## CHAPTER 1: A POLICY PERSPECTIVE ON NUMERACY AND MATHEMATICAL LITERACY ASSESSMENTS

PISA and PIAAC are policy-driven initiatives intended to provide policy makers and key stakeholders at the national and international levels with information that can inform policy-setting and planning of diverse types of social interventions and educational programs. Yet, the analytic or policy-related questions that each assessment program can answer are markedly different, given that PISA examines competencies of 15-year-old pupils who have not yet entered the world of work or adult life, while PIAAC focuses on competencies of adults in ages 16-65 years whose work experiences and personal histories can be very heterogeneous. Below we review the purposes and orientations of both assessment programs, but since these have been extensively summarised in various OECD documents, only selected points will be presented here.

PIAAC has several overarching goals (OECD, 2006, 2012). PIAAC aims to identify and measure differences within and across countries in “*literacy competencies for the information age – the interest, attitude, and ability of individuals to access, manage, integrate, and evaluate information, construct new knowledge, and communicate with others in order to participate effectively in the information age*”. Using direct assessments in three competency (cognitive) domains: *Literacy, Numeracy, and Problem-solving in technology-rich environments* (PSTRE); coupled with information collected through a set of questionnaires, PIAAC has been designed to answer many policy-related questions. For ease of presentation these policy-related questions are sketched using the three clusters summarized in Figure 1.

- The first cluster of analytic questions pertains to the middle panel in Figure 1: PIAAC is designed to provide information about distributions of competencies in the population, through direct testing of Literacy, Numeracy, and Problem-solving in technology-rich environments. The PIAAC design incorporates computer-based measures of competencies, yet also uses paper-and-pencil means for adults who are unfamiliar or uncomfortable with computers. Information is also collected about reading skills of low-performing adults.
- The second cluster of analytic questions pertains to the left-hand panel in Figure 1, i.e. understanding what are the antecedents and correlates that may help to explain the skill distribution or why and where individuals differ in actual skill levels. Such correlates include, e.g. gender, educational background, age group, or immigration status, the characteristics of jobs people hold or their literacy and numeracy-related practices at work and at home, etc.
- The last and a very important cluster, appearing in the right-hand panel in Figure 1, relates to analytic questions regarding economic and social outcomes believed to underlie both personal and societal success, such as: earnings, work status (e.g. are people employed or unemployed, how long they take to find a new job), other outcomes or processes at the individual level (e.g. use of social services, social capital such as participation in social activities), or lifespan transitions (e.g. from school to work).

**Figure 1: Simplified analytic frameworks for PIAAC and PISA**  
**What we can learn from skills surveys:**  
**School students vs. Adults**



Overall, then, PIAAC is designed to enable analyses of how actual competencies are distributed, how they are associated with a range of correlates and antecedents on the one hand, and with social and economic outcomes on the other hand. In this way, PIAAC can shed light on a range of important policy issues, such as: the adequacy of the supply of key competencies (cognitive skills), identification of groups with (relatively) low proficiencies, the extent to which the skills that individuals possess are used in their work or at home, factors that affect the acquisition and retention (or loss) of skills, and changes in skills over the lifespan.

PIAAC is further expected to enable continuity with, and links to, the two previous international adult assessments, the International Adult Literacy Survey (IALS) and the Adult Literacy & Lifeskills Study (ALL). For 19 OECD countries that participated in either IALS or ALL, it was deemed essential that PIAAC's direct assessment of cognitive skills enables capitalisation on their previous investments and provides an indication of how adults' competencies have changed since the previous measurement point(s). Thus, 60% of the literacy and numeracy tasks in PIAAC have been drawn from item pools used in ALL (and some in IALS). As a result, the conceptual framework for assessing numeracy in PIAAC is linked to the numeracy framework developed for ALL.

PISA is a policy-driven initiative, whose goals were described early on (OECD, 2003: 20) as assessing how well young people (i.e. 15-year-olds) are able to use their knowledge and skills to meet real-life challenges, rather than merely on the extent to which they have mastered a specific school curriculum. As PISA has gained momentum and worldwide attention from policy makers and educators alike, its broader goals were described by OECD in presenting the PISA 2009 results (OECD, 2010:3):

Most countries monitor students' learning and the performance of schools. But in a global economy, the yardstick for success is no longer improvement by national standards alone, but how education systems perform internationally. The OECD has taken up that challenge by developing PISA, the Programme for International Student Assessment, which evaluates the quality, equity and efficiency of school systems in some 70 countries that, together, make up nine tenths of the world economy. PISA represents a commitment by governments to monitor the outcomes of education systems regularly within an internationally agreed framework and it provides a basis for international collaboration in defining and implementing educational policies.

Wu (2010) argues that PISA's goals statement indicates that the link between achievement and curricula is not regarded as the main objective of the study. PISA adopts a "literacy" concept about the extent to which students can apply knowledge and skills. An assessment of this literacy in various subject domains is seen as having direct policy relevance for governments, though it is also intended to contribute to research and to educational practice.

The analytic questions targeted by PISA can also be examined using the three panels in Figure 1, yet there are several noteworthy differences.

1. PISA aims to examine skill/competency distributions (middle panel) and correlates and antecedents associated with the skill distributions (left panel), with a focus on a partially different set of competency domains compared to PIAAC, namely: *Reading literacy, Mathematical literacy, and Science literacy*. These are measured using a different methodology than PIAAC (see Chapters 3 and 4), given the many differences in working with students who are being tested while at school in a classroom setting, and with a much more diverse range of adults who are tested at home using a household survey methodology.
2. The correlates studied in PISA pertain to diverse factors in the school environment and home environment and to personal factors (e.g. homework preparation, TV watching habits, attitudes towards school subjects, to name just a few) that may affect student performance. These correlates are very important but markedly different from the correlates of relevance to understanding or describing performance of adults (e.g. reading practices at work, educational or occupational history, etc).
3. PISA does not examine outcomes of skills, i.e. it does not collect data about the right-hand panel in Figure 1, as its main focus is on describing variability in measured skills of students and on understanding the many relevant correlates related to both school education practices and a range of external factors. In contrast, PIAAC pays much attention to the right-hand panel, i.e., it aims to describe key outcomes (i.e. behaviors or factors of value to individuals and societies such as labor force status and employment), and understand how these are associated with actual (tested) competencies and with diverse correlates.

It follows that the goals of each program and the kinds of analytic questions that each has been designed to answer are quite different. These differences in foci in turn have affected the design of the relevant assessments, described later in Chapters 2 and 3.

## CHAPTER 2: NUMERACY AND MATHEMATICAL LITERACY: CONCEPTUAL FRAMEWORKS

Numeracy and Mathematical Literacy are related constructs, but they are not simply different names for the same underlying entity. This chapter reviews these two constructs and the trajectories along which their underlying ideas have evolved. We state at the outset that due to the breadth and depth of the literature on these and other related constructs, and given space limitations, this paper presents only selected points that sketch in broad strokes key ideas with regard to both constructs. The interested reader is advised to follow the references listed in this chapter for further information.

**Table 1: Formal definitions of key constructs**

<p><b>Document and Quantitative Literacy (IALS, International Adult Literacy Survey, 1996)</b></p> <p><i>Quantitative Literacy:</i> The knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials (such as balancing a check book, figuring out a tip, completing an order form, or determining the amount of interest on a loan).</p> <p><i>Document Literacy:</i> The knowledge and skills required to locate and use information contained in various formats (including job applications, payroll forms, transportation schedules, maps, tables, and graphics).</p> <p><b>Numeracy &amp; Numerate behavior (ALL, Adult Literacy and Lifeskills Survey, 2005)</b></p> <p><i>Numeracy</i> is the knowledge and skills required to effectively manage and respond to the mathematical demands of diverse situations.</p> <p><i>Numerate behavior</i> is observed when people manage a situation or solve a problem in a real context; it involves responding to information about mathematical ideas that may be represented in a range of ways; it requires the activation of a range of enabling knowledge, factors and processes.</p> <p><b>Numeracy &amp; Numerate behavior (PIAAC, 2012)</b></p> <p><i>Numeracy</i> is the ability to access, use, interpret, and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life.</p> <p><i>Numerate Behavior</i> involves managing a situation or solving a problem in a real context, by responding to mathematical content/information/ideas represented in multiple ways.</p> <p><b>Mathematical literacy (PISA, 2000 - 2009)</b></p> <p><i>Mathematical literacy</i> is an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen.</p> <p><b>Mathematical literacy (PISA, 2012)</b></p> <p><i>Mathematical literacy</i> is an individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens.</p>
--

## 2.1 Mathematical literacy and PISA

Mathematical literacy is a construct that seems to have a relatively short past yet a longer history. An expectation that students should have functional mathematical skills or be able to apply their mathematical knowledge to solving “everyday” or “real-life” problems has been expressed for decades in many sources around the world (e.g. NCTM, 1989; Steen, 1991; Niss, 1996; Gal, 1997; Tout, 2000; Kilpatrick, 2001; Romberg, 2001; Jablonka, 2003; Madison & Steen, 2003). Mathematical literacy first received international exposure and attention from policy makers with the publication of results from the 1995 Third International Mathematics and Science Study (TIMSS). In that cycle, TIMSS included a separate “Mathematics Literacy” test designed to “*measure how well students can use their knowledge in addressing real-world problems having a mathematics... component*” (Mullis et al., 1998) and administered to students in their *final* year of schooling (usually grade 12) in 28 countries. Results showed that many students who are about to enter adult life have difficulties in solving ‘everyday’ mathematics problems.

Mathematical literacy as a fully-fledged and stand-alone construct grabbed the attention of policy makers in full force only when OECD’s Program for International Students Assessment (PISA) was conceived around the idea that an assessment at age 15 would provide an early indication of how young adults may respond in later adult life to situations they will encounter that involve school-based knowledge and skills. In its 1999 document “Measuring student knowledge and skills: New frameworks for assessment”, OECD (1999:8) laid the foundation for the first PISA cycle in 2000:

“Indicators are designed to contribute to an understanding of the extent to which education systems...are preparing their students to become lifelong learners and to play constructive roles as citizens in society..... 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.

With the above as a foundation, the original conception of mathematical literacy developed for PISA 2000 (see Table 1) acknowledged the need to develop and hence assess students’ capacity to transfer and apply their knowledge and skills to problems that originate outside school-based learning contexts (de Lange, 2003). That definition was retained in the PISA 2003 assessment cycle, when mathematical literacy received expanded testing time and deeper coverage of sub-domains, and later used for PISA 2006 and 2009 as well. However, for PISA 2012, where the mathematical literacy domain again receives expanded coverage, the PISA Mathematical Literacy Expert Group (2010) refined and enhanced the conceptualisation of mathematical literacy, and changed some features in the assessment and reporting schemes<sup>1</sup>. (see Chapter 4 for more details, and Stacey, 2012).

For the purposes of PISA 2012, mathematical literacy has been defined as follows (and see Table 1):

Mathematical literacy is an individual’s capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens. (PISA Mathematical Literacy Expert Group, 2010)

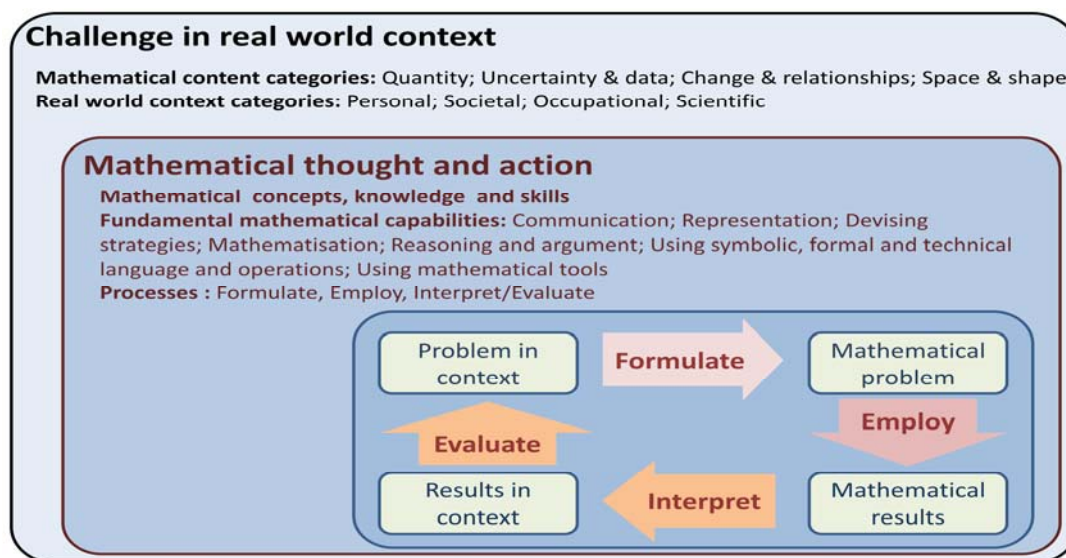
<sup>1</sup> Achieve, in cooperation with Australian Council for Educational Research (ACER), was responsible for the development and design of the 2012 mathematics framework, in consultation with the Mathematics Expert Group (MEG). Achieve also was responsible for conducting research and analyses to inform the development of the framework, including analyses of academic expectations from OECD countries and other high performing nations.

In describing changes in the mathematical literacy framework in PISA 2012, Kaye Stacey, Chair of the Mathematical Literacy Expert Group for PISA 2012 (Stacey, 2012) stated:

The new framework has clarified the definition of mathematical literacy, including emphasising the fundamental role that mathematics plays. The intention is to clarify the ideas underpinning mathematical literacy so that they can be more transparently operationalised, whilst retaining strong continuity with past assessments so that the survey outcomes provide clear evidence of trends in educational outcomes....Mathematical literacy is still seen as the understanding of mathematics central to a young person's preparedness for life in modern society, from simple everyday activities to preparing for a professional role.

The conceptualisation of Mathematical Literacy in PISA 2012, and the design of the item pools, have been in part informed by voices within the international mathematics education community that the assessment of mathematical literacy should reflect an important body of mathematical knowledge that is a part of mathematics, or that should be addressed in school education (Schoenfeld, 2004; Murphy, 2010; Lindquist, 2009). According to Stacey (2012), the revised mathematical literacy framework for PISA 2012 has aimed to dismiss the misconception that mathematical literacy is synonymous with a minimal or a low level knowledge and skills and views the competence as a continuum from low levels to high level. Core mathematical ideas underlie in a more prominent way the revised definition and conceptualisation of mathematical literacy for PISA 2012 (Stacey, 2012). One such core idea is mathematical modelling, which assumes that when individuals use mathematics and mathematical tools to solve problems set in a real world context (i.e. personal, societal, occupational, scientific), they work their way through a series of stages depicted in Figure 2.

**Figure 2. A model of mathematical literacy in practice for PISA 2012**  
(from Stacey, 2012; See an earlier version in OECD, 2010)



The processes outlined in Figure 2 of formulating, employing, interpreting, and evaluating mathematics are key components of the mathematical modelling cycle that has underpinned the ML construct in PISA since its beginnings and has been elaborated and enhanced in the description and definition of mathematical literacy for PISA 2012. These processes each draw on the problem solver's fundamental mathematical capabilities and on his or her mathematical knowledge in four overarching content areas: Quantity and

number, Uncertainty and data, Change and relationships, and Space and shape. These ideas should be compared with the PIAAC conceptualisations described below and in Chapter 4.

## 2.2 Numeracy and PIAAC

Unlike mathematical literacy, a construct that has received international prominence through its inclusion in the PISA assessment, numeracy has a much more diverse background. A detailed review of the history of conceptions of adult numeracy cannot be elaborated here for lack of space, and readers are directed to PIAAC Numeracy Expert Group (2009) and to earlier sources (e.g. Crowther Report, 1959; Coben, O'Donoghue, & FitzSimons, 2000; Hagedorn, Newlands, Blayney, & Bowles, 2003; Gal, van Groenestijn, Manly, Schmitt & Tout, 2005; Ginsburg, Manly, & Schmitt, 2006). The literature has looked at adult numeracy from an *external* perspective, i.e. what are the demand characteristics of the various real world tasks and contexts (e.g. home, community, workplace) that adults may encounter, the practices that adults develop to cope with such demands, and their [educational] ramifications (Hoyles, Wolf, Molyneux-Hodgson, & Kent, 2002; Straesser, 2003).

As summed up by the PIAAC Numeracy Expert Group (2009:9), the conceptualisation of “numeracy” in an international *assessment* context is still a challenging undertaking, because,

Like literacy, the term numeracy has multiple meanings across countries and languages. In some countries the term numeracy relates to basic skills which school children are expected to acquire as a prerequisite to learning formal mathematics at higher grades. In other countries the term numeracy encompasses a broad range of skills, knowledge and dispositions that adults should possess but it does not necessarily relate to formal schooling ... some countries do not even have a word such as numeracy.

Through a progressive process that has extended over several decades, the term “adult numeracy” is nowadays used by diverse communities to refer to people’s ability to engage with a wide range of tasks, from simple to complex, through the activation of a wide range of mathematical and statistical skills and knowledge. Numeracy refers not just to the ability to perform basic calculations but to a very wide range of skills, such as being able to measure; use and interpret statistical information; understand and use shape, design, location and direction; as well as think critically about quantitative and mathematical information, and more (Secretary’s Commission on Achieving Necessary Skills [SCANS], 1991; Tout, 2000; Tout & Schmitt, 2002; de Lange 2003; Gal et al., 2005). Further, the literature describes not only cognitive components (e.g. skills, knowledge) but also dispositional components (e.g. feelings, beliefs, self-perceptions, motivations) that contribute to adults’ numeracy. For instance, Johnston (1994:34) argues, “To be numerate is more than being able to manipulate numbers, or even being able to 'succeed' in school or university mathematics. Numeracy is a critical awareness which builds bridges between mathematics and the real world, with all its diversity. In this sense there is no particular ‘level’ of mathematics associated with [numeracy]: it is as important for an engineer to be numerate as it is for a primary school child, a parent, a car driver or a gardener. The different contexts will require different mathematics to be activated and engaged in.”

We now shift to discuss how adult numeracy has emerged as a construct included in PIAAC, and provide a brief historical account because some countries participating in PIAAC have also participated in earlier skills survey and are interested in trend data. The definitions of constructs in Table 1 show a transformation over a 20-year period, starting with an initial assessment of the useful but more restricted construct of Quantitative Literacy in national surveys in the USA and then in Canada and Australia (Wickert & Kevin, 1995). QL was proposed as part of a model developed by Irwin Kirsch and Peter Mosenthal (Kirsch, Jungblut, & Mosenthal, 1998) which viewed *literacy* as comprised of three sub-skills: Prose Literacy, Document Literacy (DL), and Quantitative Literacy (QL). Later, QL was assessed in the

world's first large-scale, international comparative assessment of adult literacy, the International Adult Literacy Survey (IALS), conducted between 1994-1996 in 19 countries, as a household survey taken by a representative sample of the adult population ages 16-65 (OECD and Statistics Canada, 2000). It is useful to note<sup>2</sup> that even though only QL seemingly involves mathematics, DL also covers some facets of what is subsumed under Numeracy in PIAAC.

The success of IALS led several governments to want to adapt the constructs and methodologies used in IALS for measuring a broader array of skills, leading to the Adult Literacy and Lifeskills Survey (ALL), which was jointly developed from 1997 by Statistics Canada and by the U.S. National Center for Education Statistics (NCES), in cooperation with the OECD. The skill domains measured in ALL included Numeracy, Document and Prose Literacy, and Problem Solving, with data collection during 2002-2008 in 11 countries and provinces (OECD and Statistics Canada, 2011). Table 1 presents the definitions for *numeracy* and for a related construct, *numerate behaviour* which were developed by the ALL Numeracy Expert Group. As can be seen, numeracy in ALL was designed to go above and beyond the Quantitative Literacy scale used in IALS where the assessment related only to applying arithmetic operations, either alone or sequentially, to numbers embedded in printed materials.

PIAAC was designed to enable its findings to be linked to previous international adult assessments with 60% of the literacy and numeracy tasks coming from the item pools used in ALL (and some in IALS). As a result, the conceptual framework for assessing numeracy in PIAAC maintains conceptual and pragmatic links to the numeracy framework developed for ALL. PIAAC's definition of numeracy closely follows the definition used in ALL and reads as follows (see PIAAC Numeracy Expert Group (2009)<sup>3</sup> and OECD 2013b for further explanations):

Numeracy is the ability to access, use, interpret, and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life.

In addition, the Numeracy Expert Group for PIAAC, based on work for ALL before it, used a more detailed specification of "numerate behaviour" which complements the above definition of numeracy:

Numerate Behaviour involves managing a situation or solving a problem in a real context, by responding to mathematical content/information/ideas represented in multiple ways.

<sup>2</sup> DL referred to the ability to comprehend and interact with texts organized in *non-linear* forms such as in forms, product labels and many other formats. Implicit in DL was also the ability to read and comprehend quantitative or statistical information in graphs, charts and tables of various kinds. QL focused on arithmetic operations on quantitative information embedded in texts either in prose text or non-linear texts (e.g. figuring out a tip as a percentage of a restaurant bill, completing an order form and calculating the cost of an order involving several items). Thus, when viewed in combination, QL and DL covered selected but important aspects of numeracy (and mathematical literacy), even though only QL seemingly involves mathematics.

<sup>3</sup> The definition of numeracy for ALL and PIAAC was designed to reflect the general conception of 'competencies' adopted by OECD's project DeSeCo (Definition and Selection of Competencies; see final report by Rychen & Salganic, 2003:8). DeSeCo has defined competency as: "[The] interest, attitude, and ability of individuals to access, manage, integrate, and evaluate information, construct new knowledge, and communicate with others in order to function effectively in the information age." One of the key changes in the definition of numeracy from ALL to PIAAC was the inclusion of "engage" in the wording, to signal that not only cognitive skills but also dispositional elements, i.e. beliefs and attitudes, are necessary for effective and active coping with numeracy situations. Previously, dispositional elements were subsumed in the ALL framework as part of "enabling processes" that underlie numerate behaviour, but including them now in the core definition itself was possible through the addition of the term "engage".



The description of *numerate behaviour* was designed to serve as a crucial link between the conceptualisation and the operationalisation of the construct of numeracy. It describes an observable behaviour and phrased in a way that can guide the development of items for the survey by putting forward four building blocks or facets (contexts, responses, content/ideas, and representations), each with several components elaborated in Table 2.

**Table 2. PIAAC's specification of numerate behaviour and its facets**  
(from PIAAC Numeracy Expert Group, 2009: 21-22)

<p><b>Numerate behaviour involves managing a situation or solving a problem...</b></p> <p><b>1. in a real context:</b></p> <ul style="list-style-type: none"> <li>- everyday life</li> <li>- work</li> <li>- societal</li> <li>- further learning</li> </ul> <p><b>2. by responding:</b></p> <ul style="list-style-type: none"> <li>- identify, locate or access</li> <li>- act upon, use: order, count, estimate, compute, measure, model</li> <li>- interpret</li> <li>- evaluate/analyse</li> <li>- communicate</li> </ul> <p><b>3. to mathematical content/ information/ ideas:</b></p> <ul style="list-style-type: none"> <li>- quantity &amp; number</li> <li>- dimension &amp; shape</li> <li>- pattern, relationships &amp; change</li> <li>- data &amp; chance</li> </ul> <p><b>4. represented in multiple ways:</b></p> <ul style="list-style-type: none"> <li>- objects &amp; pictures</li> <li>- numbers &amp; mathematical symbols</li> <li>- formulae</li> <li>- diagrams &amp; maps, graphs, tables</li> <li>- texts</li> <li>- technology-based displays</li> </ul> <p><b>Numerate behaviour is founded on the activation of several enabling factors and processes:</b></p> <ul style="list-style-type: none"> <li>- mathematical knowledge and conceptual understanding</li> <li>- adaptive reasoning and mathematical problem-solving skills</li> <li>- literacy skills</li> <li>- beliefs &amp; attitudes</li> <li>- numeracy-related practices and experience</li> <li>- context/world knowledge</li> </ul>
--

PIAAC assessment items are expected to reflect all key combinations of the four facets. The first facet in the definition of numerate behaviour shown in Table 2, *context*, is about the purpose or goal of the numeracy action, which takes place in one of four types of contexts: everyday life, work, societal, and further learning. The second facet concerns the fact that people have to respond in some way to a task in context, and describes five different *responses*: identify or locate, act upon, interpret, evaluate/analyse, and communicate about the activity or situation. The third facet is about four broad categories of *mathematical*

*content, information or ideas* embedded in the situation: quantity & number; data & chance; pattern, relationships & change; and dimension & shape. The fourth facet describes how the mathematics may appear within the context to the person who has to respond to it in some way, and describes *representations* such as concrete objects (e.g. people, buildings, cars, etc.); pictures or images of things; mathematical symbols, notations, or formulae; diagrams, charts or a maps; graphs and tables; texts; and visual displays (in print or digital form).

### **2.3 Summary**

Numeracy and mathematical literacy certainly are related constructs in terms of the core ideas that underlie them. Both refer to the ability of individuals to cope with tasks that are likely to appear in the real world and that contain mathematical or quantitative information or that require mathematical or statistical skills and knowhow. While using somewhat different terminologies, the constructs refer to quite similar building blocks and content areas. Also, both PISA and PIAAC describe mathematical literacy and numeracy as *not* synonymous with a minimal or a low level of mathematical knowledge and skills, but view the constructs as describing complex competencies lying on a continuum, i.e. individuals could be placed on a scale from low levels to high levels. However, although these are related constructs, they differ in the trajectory of development and in their analytic and reporting ecology. These and related issues will be revisited in Chapter 5 which summarises this paper, after additional details about implementation and methodology are presented in Chapters 3 and 4.

### CHAPTER 3: METHODOLOGIES IN PIAAC AND PISA AND THEIR IMPACT ON THE ASSESSMENT OF MATHEMATICAL LITERACY AND NUMERACY

This chapter briefly reviews selected aspects of the general methodologies used in PISA and PIAAC, using where possible non-technical language, in order to clarify common principles and highlight design features and key differences that affect among other things how mathematical literacy and numeracy are assessed in the two programs. Further details about the actual tests and items used to assess mathematical literacy in PISA and numeracy in PIAAC are presented in Chapter 4.

In addition, technical documents regarding all aspects of the methodologies and procedures used to implement PISA and PIAAC on the international and national levels are available from the consortia<sup>4</sup> that implement PISA or PIAAC, and from the OECD. The two consortia, in conjunction with participating countries, are responsible for the coordination and implementation of each assessment program, i.e. planning and design, scale development, translation and verification processes, training, sampling, survey operations, quality assurance, security, data handling, scoring, statistical analysis, etc. Both consortia operate under the guidance and management of the board of the participating countries (for PISA or for PIAAC) and the OECD secretariat.

#### 3.1 Overall approach and sampling and time issues

As a background, PISA and PIAAC both use large nationally representative samples of respondents chosen on the basis of standard stratified random sampling in order to collect nationally representative data that are sufficiently comparable across all participating countries. In PISA, a minimum of 4 500 completed respondents is drawn from a sample of at least 150 schools, 30 students at least randomly chosen within a school. In PIAAC, a minimum of 5 000 completed respondents nationwide are randomly sampled in the age range 16-65<sup>5</sup>. In each survey, samples have to be enlarged if a country wishes to enable separate reporting for a population subgroup of special interest to policy makers<sup>6</sup>.

The testing methodologies for both PISA and PIAAC are designed to comply with quality standards established by each consortium in line with accepted sampling practices and international testing standards. Beyond that, each program is implemented using a different assessment methodology and collects its data in a different context: The testing of 15-year-old students in PISA 2012, like in all prior PISA cycles, is accomplished in classes in a group setting. In contrast, PIAAC tests adults individually in their homes, using a standard household survey methodology that is employed by official statistics agencies across the world to collect statistics from individual citizens, such as during a census or a social survey.

For PISA purposes, students in schools can be made available for testing for several hours, assuming their school system grants permission to test them. *In PISA 2012, the test was designed for a net time of*

---

<sup>4</sup> PISA and PIAAC are each implemented by a different consortium of partner organizations (international contractors). Each country performs its own assessment process under a national program manager, with monitoring and guidance from the consortium, in order to ensure adherence to the assessment and quality standards established by the consortium and enable comparability and generalizability of results. Further details about technical aspects of the design and administration of the various waves of PISA can be found at [www.oecd.org/pisa/pisaproducts/](http://www.oecd.org/pisa/pisaproducts/).

<sup>5</sup> Age 16 and up is considered an adult in adult skills surveys and in official statistics publications in many countries, because in most countries compulsory schooling ends around age 15 or 16, and hence persons from age 16 can in principle join the workforce.

<sup>6</sup> Countries can, and some have chosen to, over-sample in order to represent subgroups of interest, e.g. separate language groups, different states or provinces, elderly from age 65 and above, persons in prisons, etc.

2.5 hours per student<sup>7</sup> (and additional time for breaks in between and for the general instructions). In contrast, testing time with adults in PIAAC is designed to be shorter, since cumulative experience in many countries by official statistics agencies has shown that adults are willing to allow into their home a representative from a statistics agency for usually no more than 1.5 hours, sometimes less. Hence *PIAAC was designed for a timeframe of approximately 60 minutes of testing time on average per respondent in addition to approximately 30 minutes for the administration of the background questionnaire*. It should be noted that respondents could in fact spend as much time as was needed to complete the assessment. The target 60 minutes of testing time was established only as a guide for test design.

The testing time available per respondent, which as stated above is much shorter for PIAAC than for PISA, inevitably dictates the *number of questions* and *type of questions* that can be presented. For example, it affects the number of background questions that can be presented, the number and proportion of multiple-choice and constructed response [open-ended] test items) and the *testing methodology*, such as the use of paper-based versus computer-based assessment, spread of items in test booklets, the use of a computer-based adaptive-testing algorithms, etc. Further, the available time per respondent has to be divided and allocated for the different elements of the assessment.

Both PISA and PIAAC use instruments for collecting two key types of information:

- a. *The actual direct assessment of cognitive skills*. This is conducted via test items, either in a paper-based or a computer-based format.
- b. *Demographic data and information about the various correlates of interest*. This was collected in PIAAC by the Background Questionnaire, and in PISA by a Student Questionnaire at the individual level, as well as by other questionnaires such as for teachers or principals. Further details about questionnaires appear in section 3.5 below.

### 3.2 Test design in PIAAC

Two key issues have informed the design of PIAAC, and their combined impact has led to a rather complex test design.

- a. *Time constraints and the need for adaptive testing*: Given constraints on testing time per respondent described above, one guideline for PIAAC was that the assessment has to be very efficient in order to extract maximum information within the time available. This has necessitated the use of a computer-based assessment as the main data collection tool (i.e. the test administrator brings a laptop into the respondent's home). The demand for efficiency regarding all cognitive measures in PIAAC required that the numeracy test is also relatively short (a maximum of 25 numeracy items are taken by an individual respondent) and is administered with a simple adaptive-testing algorithm. In order to provide the most time-efficient yet accurate estimate of the respondent's ability the respondent is not given all available numeracy items, but smaller clusters of items ("testlets"), which are chosen automatically by the computer-based testing program on the basis of the respondent's performance on a prior set of items. This in turn required that all computer-based numeracy items can be automatically scored, so that a respondent's score on one testlet can be computed automatically and serve as the basis for choosing the next testlet.

---

<sup>7</sup> Two hours for the assessment and 30 minutes for the background questionnaire.

- b. *Respondent heterogeneity and the need for a flexible assessment:* Irrespective of testing efficiency demands described above, a second issue that has guided testing design in PIAAC is that the assessment has to consider variability in respondents’ characteristics in terms of overall ability or willingness to use computers, as well as skill level. First, some respondents have little or no familiarity with computers, or feel uncomfortable to use a computer, and this had led to the use of paper-based assessments alongside computer-based assessment in PIAAC. Second, some respondents have very low skill levels or language difficulties, and hence may not be in a position to try the full cognitive test. Thus, a flexible testing design is needed that accommodates diverse types of respondents.

**Figure 3: PIAAC overall assessment design with routing**  
 (from OECD, 2012, p. 17)

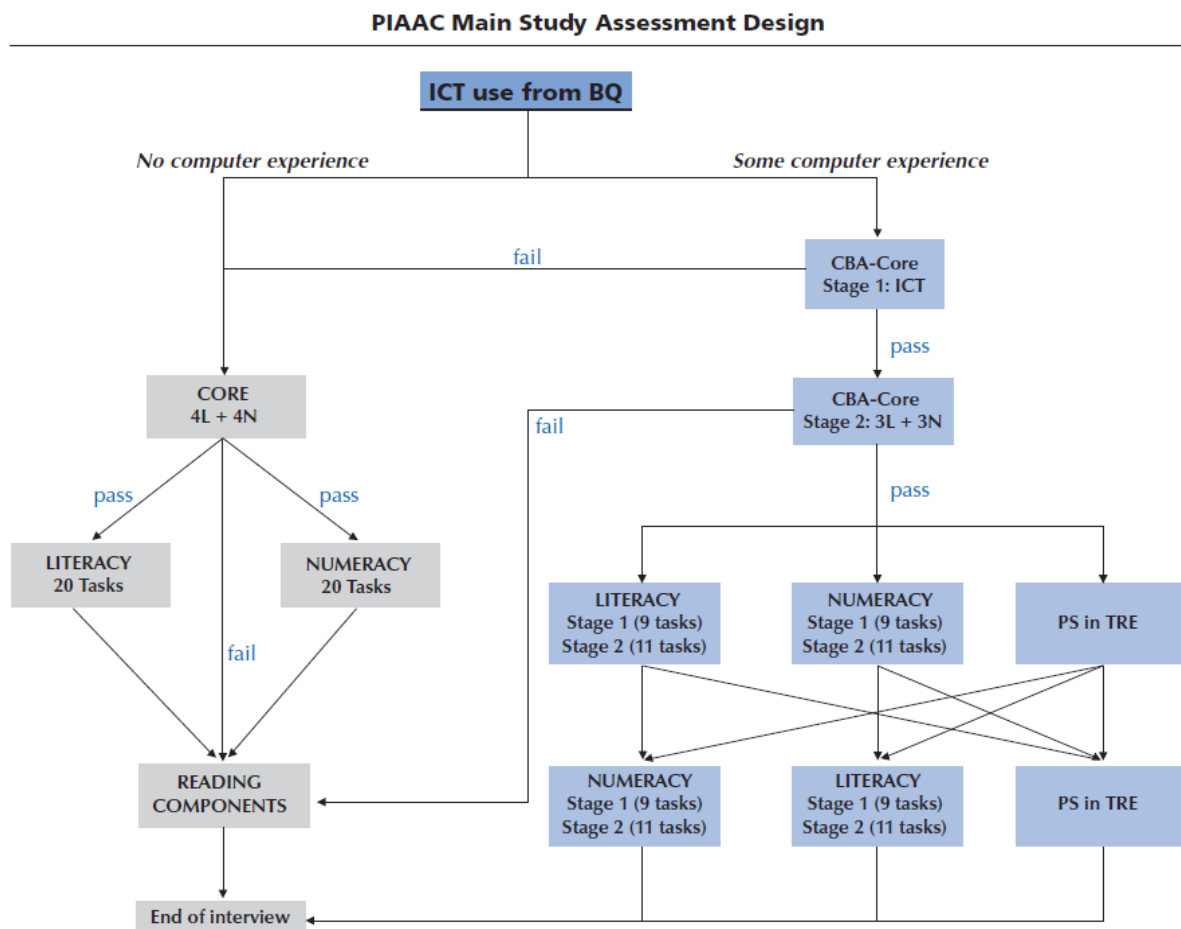


Figure 3 shows the possible pathways through the overall PIAAC assessment process, for all domains tested in PIAAC, i.e., Literacy, Numeracy, and Problem-Solving in Technology-Rich Environments. As is shown, depending on initial familiarity or skill with information and communication technologies (ICT) the respondent is routed either to a printed assessment (i.e., “pen-and-paper”; left side of the diagram), or to a computer-based assessment (right side of the diagram) done on a laptop computer which is standardized across countries and brought by the interviewer. In both pathways, respondents are then given a simple Core literacy and numeracy test comprised of several simple literacy and numeracy items. Those

failing this Core test are directed to the paper-based Reading Components assessment, designed to provide diagnostic information about specific sub-skills of low-ability individuals.

Respondents who pass the core test proceed to the full assessment under the supervision of the interviewer. The printed version of the assessment tests skills in the domains of literacy and numeracy. The majority of respondents, though, complete the computer-based assessment which can include assessments of literacy, numeracy and/or problem solving. Also, the paper-based test for numeracy includes a few additional items (not appearing in the computer-based test) requiring written constructed responses that cannot be captured and scored automatically by the computer.

PIAAC respondents have access to a hand-held calculator and can use it at any point in time during the assessment, although items have been developed so as not to demand the use of a calculator. In addition, two items require the use of a ruler (for measuring the length of objects provided as part of the assessment's attempt to simulate workplace skills), and such a ruler is provided as well.

### 3.3 Test characteristics in PISA

PISA assessments of mathematical literacy were delivered using pen-and-paper instruments from 2000 through to 2009. For PISA 2012 where mathematical literacy is the major domain, all of the participating countries (almost 70) administered the assessment in pen-and-paper based format. However, an optional computer-based assessment of mathematical literacy (CBAM) was available and this extra option was chosen by close to half of the participating countries. Both assessments will be reported against the same scale.

The PISA 2012 pen-and-paper mathematics assessment included 72 new items and 36 link items from earlier surveys to calculate trends, arranged in nine clusters of items, each representing 30 minutes of testing time. Of this total, three clusters comprise link items used in previous surveys, four 'standard' clusters comprise new material having a wide range of difficulty levels, and two 'easy' clusters devoted to items with a lower difficulty level. Each country used seven of the clusters: the three link clusters, two of the new 'standard' clusters, and either the other two 'standard' clusters or the two 'easy' clusters. The provision of "easy" and "standard" clusters allowed for a better targeting of the assessment for each of the participating countries; however, the items are scaled in such a way that a country's score will not be affected if it administers either the "easy" or additional "standard" clusters. The item clusters are placed in test booklets (forms) according to a rotated test design, with each form containing four clusters of material from the mathematics, reading and science domains. Each student does one form, representing a total testing time of 120 minutes. A combination of types of items are used in PISA, including a range of constructed response items and multiple choice items – see Chapter 4 for more details.

The optional computer-based component contains a total of four clusters of *additional* mathematics items, each cluster of 10 or 11 items representing 20 minutes of testing time. The material is arranged in a number of rotated test forms along with other material for computer delivery, with each form containing two clusters. Each student does one form, representing a total testing time of 40 minutes. The computer-based assessment<sup>8</sup> provided the opportunity to include a wider range of mathematics tools and processes that are increasingly available for solving problems, and it also provided the opportunity to assess some aspects of mathematical literacy that are not as easily assessed via traditional paper-based tests. This is elaborated further in Chapter 4.

---

<sup>8</sup> Just as pen-and-paper assessments rely on a set of basic skills for working with printed materials, computer-based assessments rely on a set of basic skills for using computers, such as knowledge of basic hardware (e.g. keyboard and mouse) and basic conventions (e.g. arrows to move forward, buttons to press to execute commands). The intention was to keep such skills to a minimal level in computer-based assessment items.

Because mathematical literacy was the major domain in PISA 2012, the PISA Mathematical Literacy Expert Group was able to review the ML framework and designed a reporting scheme that is more detailed and elaborate than in prior PISA cycles. It is intended that PISA 2012 mathematical literacy results will be reported as overall score, and scores for each of the four mathematical content categories (Quantity and number, Uncertainty and data, Change and relationships, Space and shape). In addition, a (new) set of scores is designed for each of the three key processes that are part of and underlie the mathematical modeling cycle as depicted in Figure 2, i.e. *formulate*, *employ*, and *interpret and evaluate*.

Finally, in the majority of PISA countries, respondents have access to a hand-held calculator and can use it at any point in time during the assessment, although items have been developed to be as ‘calculator neutral’ as possible. In addition, in computer based items an on-screen calculator (or a ruler when needed) are available.

### 3.4 Scoring

In PIAAC the primary cognitive assessment tool is a *computer*-based test which is automatically computer-scored due to time constraints and the need for the use of an adaptive testing algorithm. The respondents can enter responses in one of several ways, depending on the item, such as highlighting a portion of the screen, clicking on a location, choosing from a menu of options, typing in a numerical answer, etc. However, respondents who chose to take the paper-based test or were routed into it due to low ICT skills, answer by writing in test booklets which contain open-ended (constructed response) versions of all the computer-based items, as well as some additional items that require textual responses which cannot be captured and immediately scored on the computer. The written responses to open-ended questions are then manually scored by human scorers who have received extensive training, who use clearly defined Coding Guides and are closely supervised.

In PISA 2012 the primary cognitive assessment tool was a *paper*-based test. As in PIAAC, responses to paper-based tests are scored manually by human scorers against clearly defined Coding Guides and who have received extensive training and are closely monitored. In addition, students who participated in the optional PISA 2012 computer based assessment of mathematical literacy (CBAM) respond in diverse ways such as those described for PIAAC computer-based items and these are automatically scored by the computer, but unlike PIAAC, in some CBAM items respondents can also input free-form answers that are stored by the computer and later scored by human scorers.

In PIAAC, whether a respondent has taken a computer-based test (most respondents) or a paper-based test (some respondents, varying by country and age group), the responses are scored in PIAAC as correct or incorrect, i.e., there is no option for partial credit. In PISA, whilst in most items the responses are scored as correct or incorrect, in some items there is also an option for partial credit.

### 3.5 Correlates and background questionnaires

As mentioned earlier in Chapters 1 and 2, there are significant differences between PIAAC and PISA in relation to their policy goals and hence in their design. These differences affect the nature and range of the antecedents and correlates that are collected via detailed background questionnaires administered as part of both the PIAAC and PISA assessments.

A key difference between adult skills assessments such as PIAAC and school based assessment such as PISA is the rich and extensive background information that needs to be collected regarding an adult’s life history. As explained in Chapter 1, among other things PIAAC aims to describe and explain observed skill distributions in relation to various key reporting variables (e.g. age group, gender, educational background, migration status, and more) and to a wide range of other correlates (e.g. literacy and numeracy

practices at home or at work, employment history, job characteristics, to name just a few). As important, PIAAC aims to examine the connections between observed skills and a range of social and economic outcomes such as labour force status, employment and income.

Given the need to relate to a wide range of correlates and outcome variables, the PIAAC assessment includes a comprehensive Background Questionnaire (BQ) consisting of close to 300 questions across areas such as demographics, education, language, parental information, labour force participation, literacy and numeracy practices, participation in education and learning, social capital and well-being, use of technologies, and income. Further, PIAAC implements an assessment of the *Job Requirements Approach* (JRA) to gain information on the use of a broad range of generic skills in the workplace extending beyond the use of literacy, numeracy and problem solving skills. The JRA method consists of asking respondents about the importance of different types of tasks that they may be performing in their jobs as a basis for subsequently inferring the types of skills that are required. The rich array of information collected in the PIAAC BQ enables deeper research than previously possible about the acquisition of cognitive skills and their benefits and impact on an adult's life. Given the large number of items included in the PIAAC BQ and JRA instruments, in order to reduce testing time, each PIAAC respondent receives only a subset of background questions that are rotated among respondents by the computer-based testing program.

PISA measures a range of background variables that enable the analysis and reporting of mathematical literacy and other cognitive skills for important subgroups of students (e.g. by gender, language, or migration status). Students (and school principals) respond to a range of background questionnaires of around 20 to 30 minutes in length, which are central to the analysis and reporting of results in terms of a range of student and school characteristics. Questions cover, among other things, a range of variables related to students' attitudes, beliefs and emotions about mathematics. The attention to these variables is based on a research literature suggesting that the development of positive attitudes, emotions and beliefs towards mathematics is in itself a valuable outcome of schooling and predisposes students to use mathematics in their lives; and also that such variables may contribute to explaining differences in mathematical literacy.

The PISA background questionnaires also include questions about the mathematics learning environment, students' opportunity to learn mathematics at school, their interest in mathematics and their willingness to engage in it. Opportunity to Learn questions relate to student experience and familiarity with different types and styles of mathematics problems and concepts. Interest in mathematics relates to their mathematics at school and of students' perceived usefulness of mathematics outside the school classroom, and their intentions to undertake further study and/or mathematics-related careers. Willingness to engage in mathematics asks about enjoyment, confidence and (lack of) mathematics anxiety, and related beliefs of self-concept and self-efficacy.

### **3.6 Measurement scales and comparability of reported proficiencies**

Like other international assessments, both PISA and PIAAC analyse the collected data using established methods based on Item Response Theory (IRT) models. The results are reported using a continuous scale, divided into bands that are reported as levels of proficiency, and summarise performance for each participating country and subgroups within each country. Yet, there are several important technical differences between PISA and PIAAC in this regard that should be mentioned because they affect the interpretation and *limit the comparability of results* from PISA and PIAAC.

Below we provide relatively brief and non-technical explanations on three differences between PISA and PIAAC that should be taken into account by countries and researchers wishing to compare results from PISA and PIAAC with regard to mathematical literacy and numeracy: measurement scales, response probabilities (RPs) and proficiency levels. The RPs and reporting levels were determined by OECD and



the countries participating in PISA and PIAAC, based on a diverse set of considerations explained in detail in the relevant PIAAC and PISA Technical volumes. Regarding the setting of RP values for PIAAC, see also OECD (2011).

- *Measurement scales:* In PISA and PIAAC, the IRT scaling used to model performance on the cognitive measures is based on a continuous scale in the range 0-500 for PIAAC, while for PISA the scale is in the range 200-800. Such a continuous scale, regardless of its exact range, serves a dual purpose: tasks are positioned on that scale in terms of level of difficulty, and respondents are positioned on that scale in terms of a proficiency score that reflects the likelihood or probability that they can respond correctly to tasks at different levels of difficulty.
- *Response probabilities:* PISA uses a response probability value (RP) of 0.62, meaning that persons are considered as being at a certain level of proficiency if these persons are deemed likely (based on the IRT analysis of all available data) to respond correctly to tasks at that level 62% of the time. PIAAC uses a slightly higher RP of 0.67, meaning that persons are considered as being at a certain level of proficiency if they are found likely to respond correctly to tasks at that level 67% of the time. The different RPs imply that adults tested in PIAAC are held to a somewhat higher criterion compared to students tested in PISA<sup>9</sup>.
- *Reporting levels:* PIAAC reports results for five key proficiency levels on the 0-500 scale, from Level 1 to Level 5, in line with prior adult skills studies, though may also report performance below Level 1 if needed. In contrast, PISA uses six levels from Level 1 to Level 6 to describe proficiency along the 200-800 continuous scale.

To further explain the information above, we note that interpretation of what it means to be at a certain level of proficiency relates to the likelihood that a person at a certain level (other than the lowest level) would be expected to pass a test made up of items from that level. For example, if a test was made of PIAAC numeracy items classified as having a “Level 3 difficulty” based on an RP of .67, according to OECD (2011:5) a person classified as having a proficiency “at the bottom of the Level 3 band on the [domain] scale will be able to successfully undertake items with a Level 3 difficulty approximately 50% of the time, a person at the top of the level will get such items correct around 80% of the time and a person at the middle of the level will do so 67 % of the time”. Thus, a person is unlikely to be able to do well on tasks above the difficulty level associated with his or her position on the scale (but may sometimes do so because of the response probabilities). For PISA items, the same description applies but the percentages are somewhat lower, i.e., a student is expected to have a slightly lower probability (i.e. .62) to answer items correctly and still be classified as being at a certain level of proficiency.

### 3.7 Summary

The design of the PISA and PIAAC assessments has been shaped and constrained by several factors, such as the average amount of overall assessment time available per respondent which is much longer for PISA compared to PIAAC, respondent characteristics and ability or willingness to use computers, and various operational constraints. The different testing contexts, i.e., in a school (group format, more time) or home (individual format, much less time), also have many logistical ramifications, as it is more complex and expensive to sample individuals and then test them in households rather than sample schools and students within schools, and then test students in a group setting in classes. In addition, in PISA 2003 and 2012 the domain of mathematical literacy received extended coverage and hence more testing time was

---

<sup>9</sup> Both ALL and IALS used a significantly higher RP value of .80. For PIAAC this was changed by OECD and the participating countries to .67 in order to bring the performance expectations in PIAAC closer to those used in PISA. Countries that previously participated in IALS and ALL can opt to re-analyze the PIAAC data using an RP of .80 for trend analysis.

made available for relevant items and many more ML items could be included compared to PISA 2006 and PISA 2009, and to PIAAC. As a result of the combined effect of these and other factors, there are marked differences in the extent of the content coverage between the PIAAC assessment of adults' numeracy and the PISA assessment of students' mathematical literacy. Additional commonalities and differences between the actual assessment scales are described in detail in Chapter 4.

The differences in the number and range of items in PISA and PIAAC involve two other issues. First, since the primary mode of testing for PISA from 2003 to 2012 is by printed booklets, PISA is able to use constructed response (open-ended) questions that are scored by human scorers, while PIAAC's main mode of data collection is a computer-based adaptive-testing assessment system that prevents the use of many types of open-ended responses and precludes the use of human scorers given the need for automatic scoring. These features allow for a more limited range of item types in PIAAC compared to PISA. That said, both PIAAC and PISA include additional assessment mechanisms, in the form of print-based numeracy booklets in PIAAC which are given to some respondents and include constructed-response items scored by human scorers. In PISA 2012, there is an optional computer-based (CBAM) assessment for mathematical literacy that allows for a wide range of interactive items.

Despite these design-related differences, from a psychometric perspective, both PIAAC and PISA are designed using the same psychometric basis to estimate the distribution of proficiency scores in the whole population in each country. PIAAC and PISA alike combine information from both the computer-based and paper-based assessments, hence the above differences may not affect the eventual proficiency estimates as much as may first seem. Of course, given the relatively restricted number of items possible in PIAAC, only a single proficiency score will be reported in PIAAC (same as the single ML score reported in PISA 2006 and PISA 2009), whereas as explained above, for PISA 2012 the intention is to report several sub-scores as well.

In closing, we reiterate the presence of other differences between PISA and PIAAC in terms of measurement scales, response probabilities, and proficiency levels used for reporting purposes. The combined effects of these three differences, which are not specific to the assessment of mathematical literacy and numeracy, is that a comparison of reported results from PISA and PIAAC has to be done with much caution. It is not possible to directly compare the proportion of persons at a particular level in one assessment e.g. Level 2 in PISA, with the proportion at same level in the other assessment, e.g. Level 2 in PIAAC.

## CHAPTER 4: COMPARISONS OF FRAMEWORKS AND COVERAGE

The mathematical literacy or numeracy items in the PISA and PIAAC assessments are written in line with the policy goals and conceptual frameworks described earlier in Chapters 1 and 2, and taking into account the design features of each assessment described in Chapter 3. This chapter goes further to describe and compare the two surveys in terms of the [mathematical and statistical] content areas that the items cover, contexts in which the items are set, expected responses types, item formats, representations and reading demands, and the complexity schemes used to describe factors that affect item difficulty and which inform the interpretation of results from each assessment. Where relevant, in some sections details are presented in side-by-side tables in order to clarify the comparison between the assessments. Closing this chapter are some annotated examples for the types of items used in each assessment that serve to illustrate and apply the descriptions in the chapter.

Overall, comparing how the two assessments have constructed their sets of test items provides a more concrete sense for what is being measured by each assessment and for commonalities and differences between them, beyond the information offered in prior chapters. Most material in this chapter is based on an analysis of published reports and technical documents describing the development of the two frameworks (e.g. OECD 1999; OECD 2003; OECD 2009; OECD 2010; OECD 2012; PIAAC Numeracy Expert Group, 2009). However, some information in the chapter is based on the personal knowledge of the authors, who took part in the item and scale development processes in PIAAC and PISA 2012.

### 4.1 Mathematical content

Traditionally for school education, mathematics curriculum is organised around content strands (e.g. number, algebra, statistics and probability, and geometry). However, when mathematics is described for use and application outside the mathematics classroom, which is the purpose behind both numeracy in PIAAC and ML in PISA, a different approach needs to be taken, drawing on other mathematical descriptions and constructs. Both frameworks describe and develop an organisational structure for mathematical content knowledge based on how mathematical phenomena are encountered in situations in the world outside of a mathematics classroom. Accordingly, this section includes two subsections. The first, 4.1.1, describes and compares the content classifications for mathematical literacy in PISA and numeracy in PIAAC. The second, 4.1.2, discusses other issues regarding how the two frameworks relate to more traditional curriculum-based ways of describing mathematics content, and to the breadth and depth of the mathematics content included in each assessment. The reason for this separation stems in part from the difference between PISA and PIAAC noted in Chapters 1 and 2, i.e. PIAAC is expected to describe the numeracy skills of the whole spectrum of the adult population, including both people with almost no formal education as well as those with high-level skills, while PISA aims from the outset at age 15 students who have had 9 or 10 years of formal education in mathematics.

#### 4.1.1 PIAAC and PISA content classifications

Table 3 lists the details of the four key content areas covered by the numeracy assessment in PIAAC, in comparison with the four content areas of mathematical literacy covered in PISA from 2003 to 2012. As can be seen, while the PISA and PIAAC frameworks were developed by independent teams working in parallel, they use very similar descriptors for their content classifications, introducing and describing these in terms of the “big ideas” behind mathematics. They both refer, for example, to Steen (1990), who identified six broad categories: Quantity, Dimension, Pattern, Shape, Uncertainty, and Change. Interestingly, both frameworks also refer to more traditional curriculum based descriptions of their mathematical content.

PIAAC is using the same content classifications developed for ALL and used since 2000 when ALL was fielded. PISA has amended its naming and coverage of its content categories from 2000 to 2012. In the initial PISA 2000 assessment, due to restrictions on the testing time available, only two content areas, called “Big ideas” were covered: *Change and growth*; and *Space and shape*. In PISA 2003, the title was changed to “Overarching ideas” and four categories were chosen, named *Change and relationships*; *Space and shape*; *Quantity*; and *Uncertainty*. These categories were retained for PISA 2006 and 2009 but when ML became the major domain in PISA 2012 these were reviewed and refined, and became known as the “Content” areas. The only significant change was that *Uncertainty* was renamed *Uncertainty and data* to clear up some potential confusion in its overlap with the other content areas. The descriptions of the details of the content areas in PISA from 2003 to 2012, however, remained consistent and based on the same “Big ideas” conceptualisation of mathematical content.

Overall, the two frameworks appear highly consistent in terms of their descriptions and structures of the mathematical content covered in their assessments. The two assessments have specified very similar spreads across each content area (25% for each in PISA, and from 20% to 30% for each in PIAAC). PIAAC places a slightly higher emphasis on quantity and number (30%) than on pattern, relationships and change (20%) on assumption these areas are slightly less prevalent in adult life. The PISA assessment indicated an equal percentage breakdown of each classification (25%) arguably suggesting a slightly stronger emphasis on the more formal aspects of mathematical knowledge (such as algebra and formulas) which in part are subsumed under pattern, relationships and change.

**Table 3. Content descriptions of the PISA and PIAAC assessments**

<b>PISA</b>	<b>PIAAC</b>
<b>PISA 2000: Not covered</b> <b>PISA 2003 to PISA 2012: Quantity (25%)</b>	<b>Quantity and Number (30%)</b>
In PISA, the notion of <i>Quantity</i> is described as being the most pervasive and essential mathematical aspect of engaging with, and functioning in, our world. It incorporates the quantification of attributes of objects, relationships, situations, and entities in the world, understanding various representations of those quantifications, and judging interpretations and arguments based on quantity. To engage with the quantification of the world involves understanding measurements, counts, magnitudes, units, indicators, relative size, and numerical trends and patterns. Aspects of quantitative reasoning—such as number sense, multiple representations of numbers, elegance in computation, mental calculation, estimation, and assessment of reasonableness of results—are the essence of <i>mathematical literacy</i> relative to <i>Quantity</i> . Quantification is a primary method for describing and measuring a vast set of attributes of aspects of the world. It allows for the modelling of situations, for the examination of change and relationships, for the description and manipulation of space and shape, for organising and interpreting data, and for the measurement and assessment of uncertainty. Thus <i>mathematical literacy</i> in the area of <i>Quantity</i> applies knowledge of number and number operations in a wide variety of settings.	In PIAAC, <i>Quantity</i> is described as an outgrowth of people’s need to quantify the world around us, using attributes such as: numbers of features or items; costs and charges for goods and services; size (e.g. length, area, and volume); temperature, humidity, and pressure of our atmosphere; populations and growth rates of species; revenues or profits of companies, etc. <i>Number</i> is fundamental to quantification and different types of number constrain quantification in various ways: whole numbers can serve as counters or estimators; fractions, decimals and percents as expressions of greater precision, parts or comparisons; and positive and negative numbers as directional indicators. In addition to quantification, numbers are used to put things in order and as identifiers (e.g. telephone numbers or zip codes). There is also the requirement to operate on such quantities and numbers (the four main operations of +, −, ×, ÷ and others such as squaring). Facility with quantity, number, and operation on number requires a good "sense" of magnitude. Contextual judgment comes into play when deciding how precise one should be or which tool (calculator, mental math, a computer) to use. Money and time management, the ubiquitous mathematics that is part of every adult’s life, depends on a good sense of number and quantity.

<p><b>PISA 2000: Space and shape (50%)</b>  <b>PISA 2003 to 2012: Space and shape (25%)</b></p>	<p><b>Dimension and shape (25%)</b></p>
<p><i>Space and shape</i> in PISA encompasses a wide range of phenomena that are encountered everywhere in our visual world: patterns, properties of objects, positions and orientations, representations of objects, decoding and encoding of visual information, navigation, and dynamic interaction with real shapes as well as with representations. Geometry serves as an essential foundation for <i>Space and shape</i>, but the category extends beyond traditional geometry in content, meaning, and method, drawing on elements of other mathematical areas such as spatial visualisation, measurement, and algebra. For instance, shapes can change, a point can move along a locus, thus requiring a sense of function concepts. Measurement formulas are central in this area. The manipulation and interpretation of shapes in settings that call for tools ranging from dynamic geometry software to Global Positioning System (GPS) software are included.</p>	<p><i>Dimension</i> in PIAAC includes “big ideas” related to one, two, and three dimensions of “things” (using spatial and numerical descriptions), projections, lengths, perimeters, areas, planes, surfaces, location, etc. Facility with each dimension requires a sense of "benchmarks" and estimation, direct measurement and derived measurement skills. <i>Shape</i> is a category describing real images and entities that can be visualised (e.g. houses and buildings, designs in art and craft, safety signs, packaging, snowflakes, knots, crystals, shadows and plants), in both two and three dimensions. Direction and location are fundamental qualities called upon when reading, interpreting or sketching maps and diagrams. This content area requires an understanding of units and systems of measurement, both informal and standardised such as the Metric and Imperial systems.</p>
<p><b>PISA 2000: Change and growth (50%)</b>  <b>PISA 2003 to 2012: Change and relationships (25%)</b></p>	<p><b>Pattern, relationships and change (20%)</b></p>
<p>PISA describes this content area in terms of being more literate about change and relationships which involves understanding fundamental types of change and recognising when they occur in order to use suitable mathematical models to describe and predict change. Mathematically this means modelling the change and the relationships with appropriate functions and equations, as well as creating, interpreting, and translating among symbolic and graphical representations of relationships. Change and relationships is evident in such diverse settings as growth of organisms, music, the cycle of seasons, weather patterns, employment levels, and economic conditions. Aspects of the traditional mathematical content of functions and algebra, including algebraic expressions, equations and inequalities, tabular and graphical representations, are central in describing, modelling, and interpreting change phenomena.</p>	<p>In PIAAC <i>Pattern</i> is seen as a wide-ranging concept that covers patterns encountered all around us, such as those in musical forms, nature, traffic patterns ... The human capacity for analysing and identifying patterns and relationships undergirds much mathematical thinking. <i>Relationships</i> and <i>change</i> relate to the mathematics of how things in the world are associated or develop. Individual organisms grow, populations vary over time, prices fluctuate, and objects travelling speed up and slow down. Some characteristics or values can change directly in proportion or relation to another change, whilst other characteristics may change in the opposite direction or in a different way. Change and rates of change help provide a narration of the world as time marches on. The ability to generalise and to characterise relationships between variables is a crucial gateway to understanding basic economic, political or social analyses. This domain includes the ability to develop and/or use a mathematical formula between the different variables involved in a situation, alongside the need to be able to understand, use and apply proportional reasoning.</p>

<p><b>PISA 2000: Not covered</b>  <b>PISA 2003 to 2009: Uncertainty (25%)</b>  <b>PISA 2012: Uncertainty and data (25%)</b></p>	<p><b>Data and Chance (25%)</b></p>
<p>In PISA, uncertainty is described as a phenomenon at the heart of the mathematical analysis of many problem situations, and the theory of probability and statistics as well as techniques of data representation and description. <i>Uncertainty and data</i> includes recognising the place of variation in processes, having a sense of the quantification of that variation, acknowledging uncertainty and error in measurement, and knowing about chance. It also includes forming, interpreting, and evaluating conclusions drawn in situations where uncertainty is central. The presentation and interpretation of data are key concepts. There is uncertainty in scientific predictions, poll results, weather forecasts, and economic models. There is variation in manufacturing processes, test scores, and survey findings, and chance is fundamental to many recreational activities enjoyed by individuals.</p>	<p>Data and chance encompass two related but separate topics. <i>Data</i> covers “big ideas” such as variability, sampling, error, or prediction, and related statistical topics such as data collection, data displays, and graphs. Modern society demands that adults interpret and produce organisers of data such as frequency tables, pie charts, graphs and to sort out relevant from irrelevant data.</p> <p><i>Chance</i> covers “big ideas” related to probability, subjective probability, and relevant statistical methods. Few things in the world are 100% certain; thus the ability to attach a number that represents the likelihood of an instance is a valuable tool whether it has to do with the weather, the stock-market, or the decision to board a plane.</p>

#### 4.1.2 Other content related issues

There are additional issues beyond the content-related differences discussed above. These include how the two frameworks relate to more traditional curriculum-based ways of describing mathematics content, and to the breadth and depth of the mathematics content included in each assessment.

Within both the PISA and PIAAC framework documents there are also descriptions of the mathematical content phrased more in traditional school curriculum terms. These are used to help document what school based maths content is included, but are not used in the same way as the above content categories which are used for checking that the assessment covers the range of mathematical phenomenon commonly encountered in the real-world or at least the world outside the traditional school classroom. In each of the PISA Frameworks there are sections which describe and discuss the framework’s classification of content in relation to more traditional school mathematical curriculum types of descriptions. In the PIAAC framework, in the Complexity factor, Mathematical information/data, there is also quite a detailed description in terms of more school-based curriculum content.

Based on these descriptions there seems to be again very similar coverage across both PISA and PIAAC. However, there is one difference in that the PISA description incorporates a broader set of more formal content skills related to school based mathematics topics. For example, topics such as “the concept of function”; “linear and related equations and inequalities”; “similarity and congruence, and dynamic relationships involving transformation and motion of objects” are included. The PISA descriptions appear to reflect the common content descriptions of school curriculum for 15 year-olds in OECD countries. The PIAAC descriptions do not specify formal content knowledge to such an extent, but describe the content in more general terms, e.g. “formal mathematical information such as more complex formulae, knowledge of relationships between dimensions or variables”.

An examination of the item sets of both PISA and PIAAC supports the above claim and shows that PISA is interested in the ability of 15 year-olds to use and apply curriculum-based mathematical skills and knowledge in a context, whereas PIAAC is somewhat less interested in how respondents use formal mathematical skills when solving a real-life type mathematical problem. An example of this is illustrated in some of the PISA mathematical literacy items that ask respondents to be able to use information from a

real life situation and a set of data to calculate and identify specific formal characteristics of linear equations such as a gradient and the y-intercept. This type of knowledge is not assessed in PIAAC (although it does require respondents to be able to use a formula), as PIAAC generally does not require respondents to show evidence of their knowledge of the use and understanding of formal school based mathematics.

Another difference related to content coverage emerges when examining item difficulty estimates (from item performance data in the field trials), and from comparing the respective complexity schemes of the two frameworks (described in section 4.6). This difference relates to the spread of school year-level mathematics covered in the two assessments and stems from the fact noted earlier that PIAAC has to assess numeracy skills across the full breadth of an adult population, starting from a much lower level than does PISA, yet covering high-level skills as well.

Based on the content descriptions noted earlier in Table 3 and on an examination of the actual item pools in both assessments, it is quite clear that PISA in its assessment of the mathematical literacy skills focussed more closely on secondary school levels of mathematics suitable for 15 year-olds. It focuses on the beginning to upper middle years of secondary school level mathematics, with a minimal number of items that cover upper primary school level mathematics. On the other hand, PIAAC's mathematical content and difficulty level starts much lower – from early primary years through to what appears to be a similar upper level as in PISA.

To illustrate this, we can compare some of the easiest items from both assessments. In the PISA survey, items at the lower end assess, for example, numerical skills related to comparing and interpreting data in complex tables of values which include numbers into the tens and hundreds of thousands; and reading and interpreting two timetables to identify a single time satisfying a specified context-related criteria. In comparison, the easiest items in PIAAC relate to the ability to estimate or calculate the number of items in a visual image of two layers of 24 items per layer, recognising the smallest number in a one column table of numbers less than 100, or adding three numbers with a total into the low 100s. This difference is reinforced when the differences in the two complexity schemes are compared (see section 4.6 below).

## 4.2 Contexts

An important aspect of both PISA and PIAAC frameworks is that tasks or problems that have a mathematical (or statistical) content are set in a real-world context. Thus, in both assessments, items are embedded or set in a broad range of contexts that reflect the range of situations in which individuals, whether 15 year-olds or adults, have to operate in the 21<sup>st</sup> century. The two assessments use four (or five in the original PISA 2000 framework) quite similar context labels and descriptions, and have specified that items will be spread approximately equally across each of these contexts.

The first, common context in both PIAAC and PISA relates to mathematical tasks that are encountered in an individual's personal or family situations. In PIAAC these are labelled as *Everyday life* and in PISA as *Personal*. Both frameworks name situations such as shopping, games, personal health, personal transportation, sports, travel, and personal scheduling and personal finance under this classification. These descriptions are highly consistent with each other—the only apparent difference being due to the age of the two target groups—with some of the PIAAC contexts being more relevant to adults and some of the PISA contexts being more appropriate for 15 year-olds.

Another context is the one named *Societal* in PISA and *Societal or Community* in PIAAC. Despite the minor differences, these labels describe situations related to an individual's engagements with their

community or society in general. For example, both PISA and PIAAC descriptions name situations related to public policies, demographics, advertising, national statistics and economics.

A third context relates to mathematical tasks that may emerge in an individual's work or occupational situations. In PIAAC these are labelled as *Work-related* and in PISA 2012 as *Occupational*. However, in PISA 2003 (and hence 2006 and 2009) this was a combination of two separate contexts from PISA 2000—*educational* and *occupational*—which in PISA 2003 was named *educational/occupational*. The *educational* context related more to mathematical tasks and activities relevant to a classroom or school-based context, understandable given the PISA assessment was of 15 year-olds in schools. However, given that the focus of PISA is about solving authentic problems set in a real world context, in PISA 2012 the occupational context was solely used. In this context, both frameworks name and describe similar situations such as managing schedules, budgets, and project resources, quality control charts, making and recording measurements, etc. Again any difference between the two is related to the age of the two target groups with the proviso in the PISA survey that items must be accessible to 15-year-old students.

There is a fourth context in both PISA and PIAAC that on the surface look quite different. The *Further learning* context of PIAAC and the *Scientific* context of PISA are the two contexts that are the least consistent between the two surveys. In PIAAC this context is described as related to adults' needs to solving problems or dealing with tasks that may arise when participating in further study, whether for academic purposes or as part of vocational training. It is explicitly related to knowing about some of the more formal aspects of mathematics that involve symbols, rules, and formulas and to understanding some of the conventions used to apply mathematical rules and principles. In PISA there is similarity in that the *Scientific* context includes items that are named as intra-mathematical, where all the elements involved belong in the world of mathematics, not dissimilar to that in PIAAC. The PISA's *Scientific* context names areas such as weather or climate, ecology, medicine, space science, genetics, which could also be seen as having some (limited) overlap with *Societal or community* in the PIAAC framework. However, there is still considerable overlap between "further learning" in PIAAC and "Scientific" in PISA in the sense that it is within these contexts that the more formal and less context-dependent items are incorporated in both surveys.

Overall, the descriptions of item contexts in the framework descriptions for both PISA and PIAAC, coupled with an analysis of the actual items classified under each context during the item development process, imply that PISA and PIAAC refer to very similar contexts for their assessment tasks. The two assessments have also specified that items will be spread approximately equally across each of their contexts. That said, it should be noted that in actuality a few items could be classified under more than one context (a phenomenon that characterises almost all large-scale assessments and is not unique to PISA or PIAAC), and there are some appropriate variations, noted above, due to the differences in the ages of the two cohorts of respondents.

### 4.3 Responses/actions

As with the context and content classifications and descriptions, both frameworks also describe other characteristics of their assessment tasks. This includes a description of the ways in which the respondent needs to solve the task and how they might respond—a description of the processes and actions they need to take in order to answer the question posed. Table 4 lists the three **response/action** classifications covered by the numeracy assessment in PIAAC in comparison with the three related classifications of mathematical literacy covered in PISA across its different cycles. These classifications and descriptions are explained below in more detail.



**Table 4. Response classifications in PIAAC and PISA**

PISA 2012	PISA 2003-2009	PIAAC
Formulating (25%)	Reproduction (25%)	Identify, locate, or access (10%)
Employing (50%)	Connections (50%)	Act upon or use (50%)
Interpreting (25%)	Reflection (25%)	Interpret, evaluate/analyse, communicate (40%)

#### 4.3.1 PIAAC response classifications

In PIAAC it is argued that in different types of real-life situations, people may have to react with diverse types of responses, and these are classified under three key categories further explained below: *Identify, locate, or access*, *Act upon or use*, *Interpret, evaluate/analyse, and communicate*. First, the PIAAC framework argues that in virtually all situations, people have to *identify, locate or access* some mathematical information present in the task or situation confronting them that is relevant to their purpose or goal. When it exists alone, this response type often requires only low level mathematical understanding or application of simple arithmetic skills. Usually, however, this response type is subsumed or co-occurs with the other types of responses listed below.

In PIAAC the second response category, *Act upon or use*, argues that people have to perform actions on the mathematical information which can be identified in the situation, or use known mathematical procedures and rules. Acting upon or using encompasses operations such as counting, doing arithmetical calculations “in the head”, with pen and paper or with a calculator. Acting upon or using may also involve ordering or sorting, estimating, figuring out an area or volume of a certain object in an approximate way, or using various measuring devices to generate needed mathematical information of a more exact nature. Finally, acting upon may involve using (or developing) a formula which serves as a model of a situation or a process.

The third category, *Interpret, evaluate/analyse, communicate*, encompasses three separate but related responses. *Interpret* relates to the interpretation of the meaning and implications of given information of a mathematical or statistical nature. Further, in such situations, the person in the situation may need to not only interpret mathematical or statistical information but also make a judgment or create an opinion, such as about trends, changes, or differences described in a graph or in a text appearing in a newspaper article or advertisement. *Evaluate/analyse* is in part an extension of the *Interpret* response type. It accommodates responses that may be more likely in situations requiring a person to analyze a problem and in so doing evaluate the quality of the solution against some criteria or contextual demands, and if needed cycle again through the interpretation, analysis and evaluation stages.

In addition to the responses listed above, a person may have to represent and *communicate* about the mathematical information given, describe the results of one’s actions or interpretations to someone else, or explain and justify the logic of one’s analysis or evaluation.

#### 4.3.2 PISA response classifications 2003 to 2009

For PISA 2003, three categories were developed and used as an organisational tool. These were called competency clusters and were referred to as *reproduction*, *connections*, and *reflection*. They were seen as a hierarchy of competency in mathematical literacy.

*Reproduction* was described as fundamentally about reproducing practised knowledge, and matched common types of assessment tasks used in classroom tests of mathematics. It referred to knowledge of

facts, application of standard algorithms and routine procedures and manipulations. The *Connections* competency cluster was described as building on the *Reproduction* cluster and taking the problem solving process to situations that were not routine but that still are familiar. *Connections* required the application of thinking and reasoning, making connections between and applying a sequence of steps or processes from across different aspects of mathematics or different representations. The third level of the competency clusters, *Reflection*, required the activation of more complex problem solving and reasoning strategies and required some reflection and questioning of the processes to be used. This may include levels of abstraction, generalisation and modelling with unfamiliar contexts.

In PISA 2003 through to PISA 2009, the representation of items across the competency clusters was in the proportion of 1:2:1. The competency clusters were not used for reporting purposes, and in the review of the mathematical literacy framework for PISA 2012 it was decided that the competency clusters would be deleted, and were replaced with a new scheme more closely linked to both the ML definition and the mathematisation cycle at the core of the PISA ML concept and which was elaborated in Chapter 1 in Figure 2.

#### 4.3.3 PISA 2012 response classifications – the three mathematical processes

In PISA 2012 the definition of *mathematical literacy* refers to an individual's capacity to *formulate*, *employ*, and *interpret* mathematics, and this classification is used for organising the mathematical processes that describe what individuals do to connect the context of a problem with the mathematics and thus solve the problem (see also Figure 2 in Chapter 1).

The PISA 2012 process categories are:

- Formulating situations mathematically
- Employing mathematical concepts, facts, procedures, and reasoning
- Interpreting, applying and evaluating mathematical outcomes

*Formulating situations mathematically* involves identifying opportunities to apply and use mathematics—seeing that mathematics can be applied to understand or resolve some problem, or challenge, presented. It includes being able to take a situation as presented and transform it into a form amenable to mathematical treatment, providing mathematical structure and representations, identifying variables and making simplifying assumptions to help solve the problem or meet the challenge.

*Employing mathematical concepts, facts, procedures, and reasoning* involves applying mathematical reasoning and using mathematical concepts, procedures, facts and tools to derive a mathematical solution. It includes performing calculations, manipulating algebraic expressions and equations or other mathematical models, analysing information in a mathematical manner from mathematical diagrams and graphs, developing mathematical descriptions and explanations and using mathematical tools to solve problems.

*Interpreting, applying and evaluating mathematical outcomes* involves reflecting upon mathematical solutions or results and interpreting them in the context of a problem or challenge. It includes evaluating mathematical solutions or reasoning in relation to the context of the problem and determining whether the results are reasonable and make sense in the situation.

#### 4.3.4 Summary and comparison of responses/actions

A comparison of the response and process classifications across PISA and PIAAC indicates that the three sets of classifications do not match very well, even within the PISA assessments. They each try to

describe the ways in which an individual needs to solve a numeracy or ML task and the processes and actions they need to take in order to answer the question posed. Each scheme has different structures, except for some consistency between the third element of each classification, which implies some level of interpretation and reflection.

In both PIAAC and PISA 2003-2009 there is some consistency in approach, in that they are considered to be partly hierarchical and build up to more sophisticated skills or processes in their third classification. Between PIAAC and PISA 2012 there is some consistency between the second pair of categorisations in that both include performing a range of mathematical procedures and processes such as undertaking arithmetical calculations or using algebra and formulae. However, there is a difference in that the PIAAC category of *Act upon or use* also includes processes that fit more closely to some of the PISA classification of *Formulating*, i.e. the aspects relating to modelling and developing a formula. It would seem that the goal of describing the actions and processes involved within each item in the PIAAC and the multiple PISA assessments has taken different approaches. It will be interesting to observe if the three PISA process categories can be used for reporting purposes in PISA 2012 and whether it will provide a research base for further analysis of the processes used in solving a numeracy or mathematical literacy problem.

#### 4.4 Item formats

An important aspect in designing assessments is item format. Further, the ability of an assessment to actually capture, evaluate, and score responses depends on the delivery mode, operational issues and technical aspects of that assessment. There is a difference in the response style and interactivity available for items when delivered in a computer based environment compared to a traditional pen-and-paper based environment.

The traditional divide in item format is between selected-response (sometimes called forced-choice or multiple choice) format versus a constructed-response format. Selected-response items require the choice of one or more responses from a number of response options. Responses to such questions can usually be automatically processed and scored when presented on a computer or by scanning a response sheet. Within constructed response items there are open constructed-response or closed constructed-response. This is also complicated further by whether the assessment is based on a pen-and-paper based assessment or a computer-based assessment. In computer-based assessment the issue of marking open constructed-responses is more complex and requires online or computer-based systems for markers to access the respondent responses.

Open constructed-response items require respondents to communicate in their own words the answers to tasks or questions given as part of the assessment. Such items also may ask the student to show the steps taken or to explain how the answer was reached. These items require trained experts to manually code responses. In contrast, closed constructed-response items provide a more structured setting for answering, and they aim to produce a response that can be easily judged to be either correct or incorrect. The most frequent closed constructed-responses in mathematics or numeracy tests are required in “fill-in-the-blank” tasks in which the numerical responses are deliberately constrained. Often responses to questions of this type can be keyed into data capture software, and coded automatically, but some must be manually coded by trained experts.

##### 4.4.1 PISA item formats

PISA 2003 to 2009 has developed and used a range of item formats for assessing mathematical literacy in standard printed booklets, and these item formats were combined with new item formats developed for PISA computer-based assessments in 2009 and 2012. The PISA 2012 item classification can

thus be and has been retrospectively applied to PISA 2000-2009 mathematical literacy link items and serves as a more up-to-date and comprehensive description of item types. The item format categories described for PISA 2012 are listed below (and these are used as a basis for comparing PISA and PIAAC item types in section 4.4.3):

- CRE – Constructed Response Expert – items that require expert coding
- CRA – Constructed Response Auto-coded – items that can be auto coded (key in the response, then apply algorithm)
- CRM – Constructed Response Manual – that have a very limited range of full credit responses but are best coded manually
- SRS – Selected Response Simple – Simple Multiple Choice – all can be auto coded. These include any item where there is ONE correct response that the student selects. This includes both radio buttons and a single drop down menu where there is a unique correct response
- SRC – Selected Response Complex – Complex Multiple Choice – all can be auto coded
- SRV – Selected Response Variations (match opinion, Likert). These include any item in which the student selects a response that is NOT SRS or SRC. This includes drop down menu items where either a) there is more than one drop down menu; b) there is more than one possible correct response; or c) where more than one choice may be made. For example, select the best two responses from the following list

While the above item format types cover both paper-based and computer-based items, it is useful to further elaborate about the nature of computer-based item formats, which appear both in PISA and PIAAC. There is a difference in the style and interactivity available for computer-based items compared with paper-based items. The computer provides a range of opportunities for designers to write test items that are interactive, authentic and engaging, e.g. drag-and-drop items; the use of hot spots on an image to allow students to respond to more items non-verbally; the use of animations including representations of three-dimensional objects that can be manipulated; to present students with real-world data (such as a large, sortable dataset); or the use of colour and graphics to make the assessment more engaging. By design, not all computer-based items use new item formats, which might be helpful in monitoring the (positive or negative) impact that new item formats have on performance. Other benefits and advantages of computer-based test items are that the computer-based testing system can also collect data and keep a log about what the respondent did, such as the time taken to solve an item, number of clicks, processes followed, or the final state reached before the respondent moved to the next item.

With the above in mind, a computer-based item type classification scheme was also developed (PISA Mathematics Expert Group, 2011) to classify the types of items that were utilized in the platform available for the delivery of PISA in 2012. The categories described were:

- Animation, and/or manipulation
- Automatic calculation, where menial/calculation dependent work can be automated “behind the scenes” to enable or support assessment of other mathematical skills and understanding
- Drawing, spatial, visual cues and/or responses
- Interactive graphing
- Simulation of computer-based applications, for example, spreadsheets
- Simulation of web-based applications or contexts
- Items that require no specific computer-based application and could have been delivered on paper

Lastly, it should be noted that the design of computer-based item formats for PISA 2012 was informed by the possibility to use in PISA 2012 trained, expert coders to mark and code both Constructed Response Expert and Constructed Response Manual type items for both its paper-based and computer-based assessments. This is because in PISA 2012, responses were stored by the computer and could be scored at a later point in time by the expert coders. However, the use of expert coders was not possible for the computer-based assessment in PIAAC, as explained below, and hence certain item types that exist in PISA 2012 were not developed in PIAAC.

#### *4.4.2 PIAAC item formats*

PIAAC's design specification (see Chapter 2) imposed a number of restrictions on possible item formats for numeracy assessment in PIAAC:

- The need to provide the two optional pathways for respondents—the computer-based path (for most respondents) or the paper-based path (for respondents unable or unwilling to use a computer)
- the need for immediate scoring of responses, given the adaptive testing process necessary for efficient ability estimation
- the need to minimise the testing time per respondent which is typical for large scale household surveys. [Note that there was no time limit instituted for respondents in PIAAC – respondents could in fact spend as much time as was needed to complete the assessment]

In PIAAC, where most respondents were assessed in a computer-based format, these realities necessitated the use of short tasks to which the responses could be automatically scored, and excluded the use of extended response problem tasks in the computer-based assessments. The PIAAC computer-based system does allow respondents to provide an answer in several different modes, such as: numeric entry, clicking on an area of the screen, or choosing from pull-down menus. In PIAAC, tasks requiring communication-based responses, such as when adults have to explain interpretations of given information, or describe their evaluation or analysis of a situation, could not be used in the direct assessment on the computer platform, though open constructed-response items were included in the paper-based route. In an attempt to assess communication skills on a computer without the use of manual scoring, a few of the PIAAC numeracy questions ask respondents to provide an explanation for a response by choosing from predesigned encapsulated texts that appear on-screen, so as to simulate the way a person provides a justification for an answer in real life. However, such solutions have their own limitations.

#### *4.4.3 Summary and comparison of item formats*

Table 5 compares the range of item formats utilised in PIAAC and PISA based on the categories developed for PISA 2012 described above. Note that the classification under “PISA paper-based” in Table 5 applies to all PISA assessments from 2000 through to 2012.

Table 5. Item formats in the PISA and PIAAC assessments

<b>PISA – computer (2012)</b>	<b>PISA – Paper-based (2000 – 2012)</b>	<b>PIAAC- computer</b>	<b>PIAAC– Paper-based</b>	<b>Item format</b>
<b>Item scoring type:</b>				
<b>Yes</b>	<b>Yes</b>	<b>No</b>	<b>Yes, limited number</b>	Constructed Response Expert
<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>No</b>	Constructed Response Auto-coded
<b>Yes</b>	<b>Yes</b>	<b>No</b>	<b>Yes</b>	Constructed Response Manual
<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	Selected Response Simple – Simple Multiple Choice
<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	SRC –Selected Response Complex
<b>Yes</b>	<b>No</b>	<b>Yes</b>	<b>Yes</b>	Selected Response Variations
<b>PISA 2012 - computer</b>		<b>PIAAC- computer</b>		<b>Computer based item type:</b>
<b>Yes</b>		<b>No</b>		Animation
<b>Yes</b>		<b>No</b>		Automatic calculation
<b>Yes</b>		<b>No</b>		Drawing, spatial, visual cues
<b>Yes</b>		<b>Yes</b>		Interactive graphing
<b>Yes</b>		<b>No</b>		Simulation of computer-based applications
<b>Yes</b>		<b>Yes</b>		Simulation of web-based applications or contexts
<b>Limited number</b>		<b>Yes – majority of the items</b>		No specific CBAM application and could have been delivered on paper.

As shown in Table 5, the PIAAC computer-based assessment is considerably more limited in terms of its range of item types and responses available compared to PISA. This partly stems from the fact that PISA uses trained, expert coders to mark and code both Constructed Response Expert and Constructed Response Manual type items for both its paper-based and computer-based assessments, whereas PIAAC does not due to its operational restrictions and need for automatic coding, explained earlier. Other reasons relate to the time restrictions and the need to run equivalent parallel paper-based and computer-based assessments in PIAAC, which do not exist as such in PISA. Further, the development of the PIAAC computer-based platform occurred earlier than that for PISA, hence there was more opportunity for the PISA computer-based items to be more interactive.

Given the above, PISA has the ability to assess more extensive, written responses by utilising expert marking of extended response type items in the paper-based tests which are the primary mode of delivery of assessment in PISA. This enables PISA to offer potentially richer descriptions of students' abilities to reason and communicate their mathematical skills and knowledge. In contrast, PIAAC's relies mainly on computer-based assessments where the range of item types is more restricted and there is very limited

possibility for open constructed responses. That said, although there are several differences in the items types in the two assessments, eventually PIAAC combines the information from the computer-based and paper-based portions of its assessment (see Chapter 2) in order to generate population level estimates for skills distributions. This process to some extent compensates for the limitations imposed on the PIAAC computer-based assessment.

#### **4.5 Representations, reading demands, and authenticity-related issues**

Respondents' performance can be influenced by a range of additional factors beyond those noted earlier, including but not limited to three issues noted in this subsection: how information is represented within assessment tasks, the amount and difficulty level of the text components of items, and how it is perceived by the respondents because of its degree of familiarity or authenticity.

##### *4.5.1 Representations*

In both the PIAAC and PISA assessments, the mathematical information is represented in the same range of forms and styles. This is described in both frameworks. In PIAAC it states that mathematical information in a real-life situation may be available or represented in many forms including: concrete objects to be counted (e.g. people, buildings, cars, etc.) or as pictures of such things; through symbolic notation (e.g. numerals, letters, and operation or relationship signs); formulae; a diagram or chart; graphs and tables. It goes on to describe that mathematical information may also be embedded in various types of texts, either in prose or in documents with specific formats with words or phrases that carry mathematical meaning, or is expressed in notations or symbols (e.g. numbers, plus or minus signs, symbols for units of measure, etc.). In PISA, it similarly states that representations can include graphs, tables, diagrams, pictures, equations, formulae, textual descriptions, and concrete materials.

##### *4.5.2 Reading demands*

The level of reading required to successfully engage with an item is considered very carefully in item development and selection for both PIAAC and PISA. A goal in the item development process for both programs is to make the wording of items as simple and direct as possible and this is stated clearly in the both the PISA and PIAAC frameworks, including in the initial PISA ML Frameworks. Care is also taken to avoid item contexts that would create a cultural bias, and choices are checked through a number of processes including with national teams and translation experts. Translation of the items into many languages is conducted very carefully, with extensive back-translation and other protocols.

For PISA 2012 where ML was the major domain again, and new item development was undertaken, items were developed with a focus on the accessibility of the items by reducing the reading demands and also by the use of photos, images and illustrations. Indeed, in an independent review of the new PISA 2012 mathematical literacy items conducted by Achieve and in a validation process undertaken by a team of international experts, the report concluded that the PISA 2012 item pool had improved over previous cycles in relation to reading demands and that there was less unnecessary information.

Likewise, PIAAC items were also deliberately written to minimise the reading demands and to distinguish performance in numeracy more clearly from the other measures of literacy. This was achieved, as with PISA, by reducing the complexity of the text, alongside minimising the amount of text and the use of supporting photos, images and illustrations. At the same time, the development of numeracy items for PIAAC also took into account that text is an inevitable aspect of adults' life and hence may have a role in some numeracy tasks thus should not be eliminated in ways that will reduce task realism or authenticity.

A comparison of items in both survey instruments indicates that the majority of items are very similar regarding text-related demands. However, some of the PISA items do appear to have a higher complexity of text which can be explained by the differences in the mathematical content as described above in section 4.2. This is because some PISA items, more often than PIAAC items, include some formal and complex mathematical terminology and symbols, or use text to describe details of a situation in order to make it familiar to school students. As well, some PIAAC items have very little text at all, which is possible because of the need in PIAAC for items that can assess lower level mathematical skills which may be needed in tasks which contain little text.

#### 4.5.3 *Authenticity of items and tasks*

Both PIAAC and PISA items are developed on the basis of finding situations and tasks from across different countries that are couched in the four context classes described earlier, and based on authentic stimuli (or when needed due to technical reasons such as copyright, or reading difficulty or accessibility reasons, stimuli are designed to simulate authentic tasks). However, there are differences between the way adults and 15 year-olds may approach some such tasks, given the diversity in their experiential backgrounds, their respective distances from schooling, their literacy and numeracy practices, or attitudes and dispositions. Issues related to confidence and presence of mathematics anxiety/fright, and related beliefs such as self-concept or self-efficacy, are factors that could also impact on how respondents, whether 15 year-olds or adults, may react to assessment tasks. Because many adults may not remember school-based notations or symbols, the design of PIAAC items took into account not just the need to retain authenticity but also to reduce the use of task elements (e.g. formal notations and ‘school-like’ appearance) which could produce negative reactions that might lower actual cognitive performance.

It can be assumed, based on the research literature, that many adults have personal experiences and ways of coping with everyday situations which are different than those of school-age students. Hence, the types of responses envisioned of adults tested in PIAAC, and the explanations for underlying enabling factors in the numeracy framework for PIAAC, are not couched in a school-oriented “mathematical problem-solving” culture as much as in PISA. PISA relates to school-based populations, and while it is interested in students’ performance on real life problems, an underlying assumption is that the performance is to be based on skills and dispositions acquired in a schooling context. As a result, descriptions of students’ desired actions or underlying cognitive processes can be couched in a school-based environment. Indeed, an examination of PISA items shows that some items expect the understanding and use of formal symbolism that reflect an expectation for formal knowledge of what was taught in schools, yet such knowledge is less (or not) available to adults who have been out of formal school environment for years. Because many adults may not remember school-based notations, PIAAC items are less likely to include such style of representations, although some items do at the higher levels. As a result, some PIAAC items may seem to school-oriented math educators or to mathematicians to be less “mathematical”, even though they may not differ from PISA items in actual demands in terms of the mathematisation process or underlying reasoning.

The extent to which item authenticity influences motivation and engagement with assessment tasks by school-age students compared to adults has not been the subject of extensive research. That said, there is cumulative experience about such issues among educators working with adult mathematics learners. Certainly issues regarding item authenticity are an area of potential rich research to which results from the numeracy and ML assessments in PISA and PIAAC could contribute.



## 4.6 Item complexity factors

Each assessment framework also has at its core a scheme describing the various factors affecting item complexity. These schemes are used internally by item development teams and expert groups for various purposes, i.e. to inform item design, to evaluate items chosen for inclusion in the final assessment, and to inform the descriptions or interpretations attached to different performance levels on the assessments. This section will offer a comparative analysis of the item complexity schemes for PISA and PIAAC, which has not been attempted before, to shed additional light on what is actually being measured by each assessment.

### 4.6.1 PIAAC complexity scheme

PIAAC is using a complexity scheme originally developed for ALL in order to predict the difficulty or complexity of a numeracy assessment task. A scheme of five factors was developed that attempted to account for the difficulty of different tasks, enabling an explanation of observed performance in terms of underlying cognitive processes or factors. Table 6 summarises the five factors, and shows that two of them relate to textual aspects of numeracy tasks, and three relate to mathematical aspects of tasks. Additional information about each factor appears in Table 7.

**Table 6. A summary of PIAAC complexity factors**

Range	Factor	Aspects
Obvious/explicit to embedded/hidden	1. Type of match/problem transparency	<b>Textual aspects</b>
No distractors to several distractors	2. Plausibility of distractors	
Concrete/simple to abstract/complex	3. Complexity of Mathematical information/data	<b>Mathematical aspects</b>
Simple to complex	4. Type of operation/skill	
One to many	5. Expected number of operations	

These five factors were used by the ALL and PIAAC numeracy expert groups to estimate, separately and in interaction, the difficulty level of numeracy tasks. For each of these factors a detailed description was developed against a scoring system in the range from 1 through to 5, for a total difficulty score in the range 5 to 19. The five factors are described more fully below (details in the Annex of Gal et al., 2005):

**Table 7. Details of the PIAAC complexity factors**

<p><b><i>Type of Match/Problem Transparency.</i></b> This is a combination of the factor of Problem Transparency outlined above, and of an IALS factor called Type of Match. Problem Transparency is a function of how well the mathematical information and tasks are specified and includes aspects such as how apparently the procedure is set out, how explicitly the values are stated, etc. Type of Match refers to the process that a respondent has to use to relate the requested action in the question to the information in the task or text, which can range from a simple action of locating or matching to more complex actions that require the respondent to perform a number of searches through the information given. This measure of complexity for a numeracy task incorporates the degree of text embeddedness of the mathematical information.</p>
--

**Plausibility of distractors.** This variable is literacy related, even though it can involve mathematical components. In general, literacy tasks are easiest to process when there are no plausible distractors in the text, that is, there is no other information in the text that meets any of the requirements of the task. At higher levels of difficulty, tasks can involve irrelevant information both within the question as well as within the text. In terms of mathematical information, a low level of plausible distractors would mean that no other mathematical information was present apart from that requested, making the numbers or data required easy to identify. At a higher level, there may be either some other mathematical information in the task (or its text) that could be a distractor, or the mathematical information given or requested could occur in more than one place. A higher level of complexity could also mean that outside information (e.g. the knowledge of a formula) may be needed to answer the question.

**Complexity of Mathematical Information.** Some situations present a person with simple mathematical information, such as concrete objects (to be counted), simple whole numbers, or simple shapes or graphs. At lower skill levels, the information will be more familiar, whereas at higher levels, the information may be less familiar. Situations will be more difficult to manage if they involve more abstract or complex information, such as very large or very small numbers, unfamiliar decimals or percents, information about rates, or dense visual information, as in a diagram or complex table.

**Type of Operation/Skill.** Some situations require simple operations, such as addition or subtraction, or simple measurement (e.g. finding the length of a shelf), or recognition of shape. These are usually easier to analyze mathematically than situations that require multiplication or division, and than situations that require using exponents. While the difficulty of recognizing and carrying out the operation implied by a situation (be it additive, multiplicative, etc.) has direct bearing on task complexity, there may be exceptions that occur when alternative approaches are obvious. There are some tasks that combine both interpretive and generative skills and may involve a deeper conceptual understanding than merely carrying out a procedure. Other more complex tasks may involve an explanation of one's reasoning. The interpretation of information appearing in graphs, for example, becomes more complex if comparisons, conjecturing, or "reading beyond the information given" is required.

**Expected Number of Operations.** Tasks that require acting upon the mathematical information given may call for one application (step) of an operation, or for one action (e.g. literal reading of information in a table, or measurement). More complex tasks will demand more than one operation, which may be the same or similar to one another, such as the steps involved in multiple passes on the data or text. Still more complex tasks are those that involve the integration of several different operations.

The PIAAC complexity scheme reinforces the PIAAC understanding and description of numeracy as having more than just a focus on mathematics—it has a literacy component—the ability to read and interpret and extract the mathematics from the real world context and then also take the mathematics back into the real world and interpret, reflect and describe the mathematics. This component built on the knowledge of earlier models of Document Literacy and Quantitative Literacy (Kirsch, Jungblut, & Mosenthal, 1998). These models have described in theoretical terms the cognitive operations involved in coping with tasks requiring reading of documents such graphs, charts and tables, or conducting arithmetic operations on quantitative information embedded in texts, such as:

- *reading* information from various sources;
- *cycling* through various parts of diverse texts or displays;
- *integrating* information from several locations (e.g. across two graphs);
- *generating* new information (e.g. finding the difference between percentages in different parts of a table or between bars in a graph); and
- *making inferences*, quite often in the presence of irrelevant or distracting information.

The other three mathematical factors of the PIAAC complexity scheme focus on and describe three aspects that can make the mathematical operations and tasks to be utilised and applied in the context more difficult or not.

#### 4.6.2 PISA Fundamental Mathematical Capabilities and item complexity factors

Independently, of the efforts described above regarding complexity factors in ALL and PIAAC, over the years the PISA Mathematical Literacy Expert Group has also developed a description of a set of *fundamental mathematical capabilities*, i.e. factors that underpin mathematical literacy in practice. The work of Mogens Niss and his Danish colleagues (Niss, 2003 and Niss & Jensen, 2002) identified eight capabilities—referred to as “competencies” by Niss and in the 2003 PISA framework—that are instrumental to mathematical behaviour. The PISA 2012 framework uses a modified formulation of this set of capabilities, which condenses the number from eight to seven, and has been further refined and developed since the publication of the PISA 2012 framework document in 2010. Table 8 presents the description of the seven fundamental mathematical capabilities in the latest PISA mathematical literacy framework.

As with PIAAC, for PISA items can be scored against these factors using a scoring system in the range from 0 through to 3. Six of these factors and their scoring system (excluding the last factor, *Using mathematical tools*) were used by the PISA test developers and the Mathematics Expert Group to estimate the difficulty level of the PISA mathematical literacy tasks. The PISA complexity scheme is clearly designed to help expand, describe and address the mathematical modelling or mathematisation cycle which underlies all of the PISA ML frameworks. As such its seven factors illustrate a range of components that impact on that cycle, giving it a different flavour and focus to that of PIAAC, with its interest and focus on the two key aspects of using literacy skills alongside mathematical skills to unpack and solve a numeracy problem.

**Table 8. PISA 2012 –fundamental mathematical capabilities that underlie item complexity**

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

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

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

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

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

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

**Using mathematical tools:** The final mathematical capability that underpins *mathematical literacy* in practice is *Using mathematical tools*. Mathematical tools encompass physical tools such as measuring instruments, as well as calculators and computer-based tools that are becoming more widely available. This ability involves knowing about and being able to make use of various tools that may assist mathematical activity, and knowing about the limitations of such tools. Mathematical tools can also have an important role in communicating results. Previously it has been possible to include the use of tools in paper-based PISA surveys in only a very minor way. The optional computer-based component of the PISA 2012 mathematics assessment will provide more opportunities for students to use mathematical tools and to include observations about the way tools are used as part of the assessment.

#### 4.6.3 Summary and comparison of item complexity factors.

The PISA and PIAAC complexity schemes appear to be quite different, and each reflects a different development trajectory. The PISA scheme builds on its mathematical modelling or mathematisation cycle, while PIAAC builds on the evolution of the numeracy construct from a Quantitative Literacy framework and hence distinguishes between the influence of factors related to textual and mathematical aspects of numeracy tasks. One specific area where the two complexity schemes may converge is in how they describe the role of literacy and text. As mentioned above, the PIAAC complexity scheme names two literacy related factors, whereas the PISA set of factors appears not to incorporate this aspect as explicitly in the names of the factors. However, the fundamental mathematical capability called *Communication* describes four levels of the capability using statements such as “Identify and extract relevant information ... or cycle within the text or between the text and other related representations.” These ideas are quite comparable to PIAAC’s literacy based factors *Plausibility of distractors* and *Type of match/problem transparency*. Thus, the two schemes are not as different as may at first appear. It would be interesting (and valuable) to apply both schemes to the same set of items to see how they correlate, and whether there could be ways of integrating aspects the two schemes to improve their use in predicting the difficulty level of mathematical literacy or numeracy tasks more accurately.

#### 4.7 Sample items

Following are four examples of publicly available items from PIAAC, ALL and PISA that help illustrate differences and similarities between the types and styles of items in both the PISA and PIAAC assessments.

The PIAAC item in Figure 4 satisfies criteria for both assessments and would fit within the contexts of *Societal or community* in PIAAC or *Societal* in PISA; the content areas of *Data and chance* or *Uncertainty and data* respectively; and into the response types of *Interpret, evaluate* and *Interpreting, applying and evaluating* in PIAAC and PISA respectively. The item response type is *Selected Response Variations*; and it is an item that has no computer-based interactivity and could be delivered in a paper-based format. It has a small amount of text, but the language in the stimulus and question requires careful reading and interpretation.

Figure 4: PIAAC item



Look at the graph about the number of births. Click to answer the question below.

During which period(s) was there a decline in the number of births? Click all that apply.

- 1957 - 1967
- 1967 - 1977
- 1977 - 1987
- 1987 - 1997
- 1997 - 2007

The following graph shows the number of births in the United States from 1957 to 2007. Data are presented every 10 years.



Year	Number of Births
1957	4,300,000
1967	3,520,959
1977	3,326,632
1987	3,809,394
1997	3,880,894
2007	4,315,000


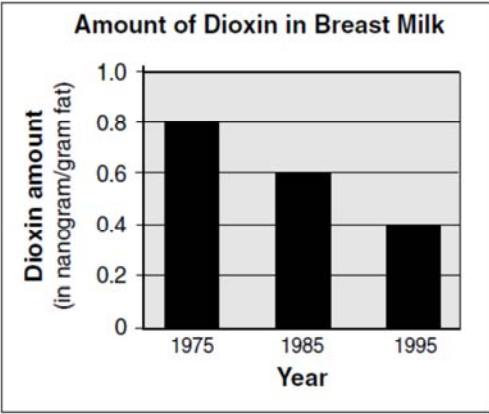


Figure 5: ALL item similar to PIAAC items

## Is breast milk safe?

Since the 1970s, scientists have been worried about the amount of Dioxin, a toxin in fish caught in the Baltic sea. Dioxin tends to accumulate in breast milk and can harm newborn babies.

The diagram shows the amount of Dioxin in the breast milk of North European women, as found in studies done from 1975 to 1995.



Year	Dioxin amount (in nanogram/gram fat)
1975	0.8
1985	0.6
1995	0.4

**Refer to the article titled ‘Is breast milk safe?’ to answer the question:** Compare the percent of change in the Dioxin level from 1975 to 1985 with the percent of change from 1985 to 1995. Which percent of change is larger? Explain your answer.

The ALL item in Figure 5 satisfies criteria for both assessments and would fit within the contexts of *Societal or community* in PIAAC or *Societal* in PISA; the content areas of *Pattern, relationship and change* or *Change and relationship* respectively; and into the response types of *Interpret, evaluate* and *Interpreting, applying and evaluating* in PIAAC and PISA respectively. The item response type is *Constructed Response Expert*; and it is an item that has no computer-based interactivity. As it is a *Constructed Response Expert* type of item it could **not** appear in the computer-based delivery of PIAAC. It has a small amount of text, but the language in the stimulus and question requires careful reading and interpretation and requires a high level of mathematical understanding.

Figure 6: PISA item

**CAR COST CALCULATOR**

To promote train travel, the Zedtown Transportation Service is distributing a car cost calculator.

The calculator compares costs for car travel from home to work and back with the cost of a monthly train ticket worth 98 zeds.

You can use the calculator by clicking and dragging the car to set the distance from home to work. The window CAR COSTS shows the monthly cost of going to work and back by car.

**DISTANCE**  
1 km  
Home to work

**CAR COSTS**  
116 zeds  
Monthly cost going to work and back by car

Zedtown Transportation Service

**MONTHLY TRAIN TICKET**  
98 zeds

**ZEDTOWN**

**Question 3: CAR COST CALCULATOR** CM013Q03

The formula for working out the car travel costs needs to take into account more than just the petrol costs. The Zedtown Transportation Service adds an additional value of  $b$  zeds per month to the monthly petrol costs to allow for other car costs such as insurance and registration.

The formula they use to work out the costs is:  $C = 6d + b$

$C$  is the total cost in zeds,  $d$  is the distance to work in kilometres, and  $b$  is the additional non-petrol costs in zeds per month.

Use the car cost calculator to help you calculate the value of  $b$ .

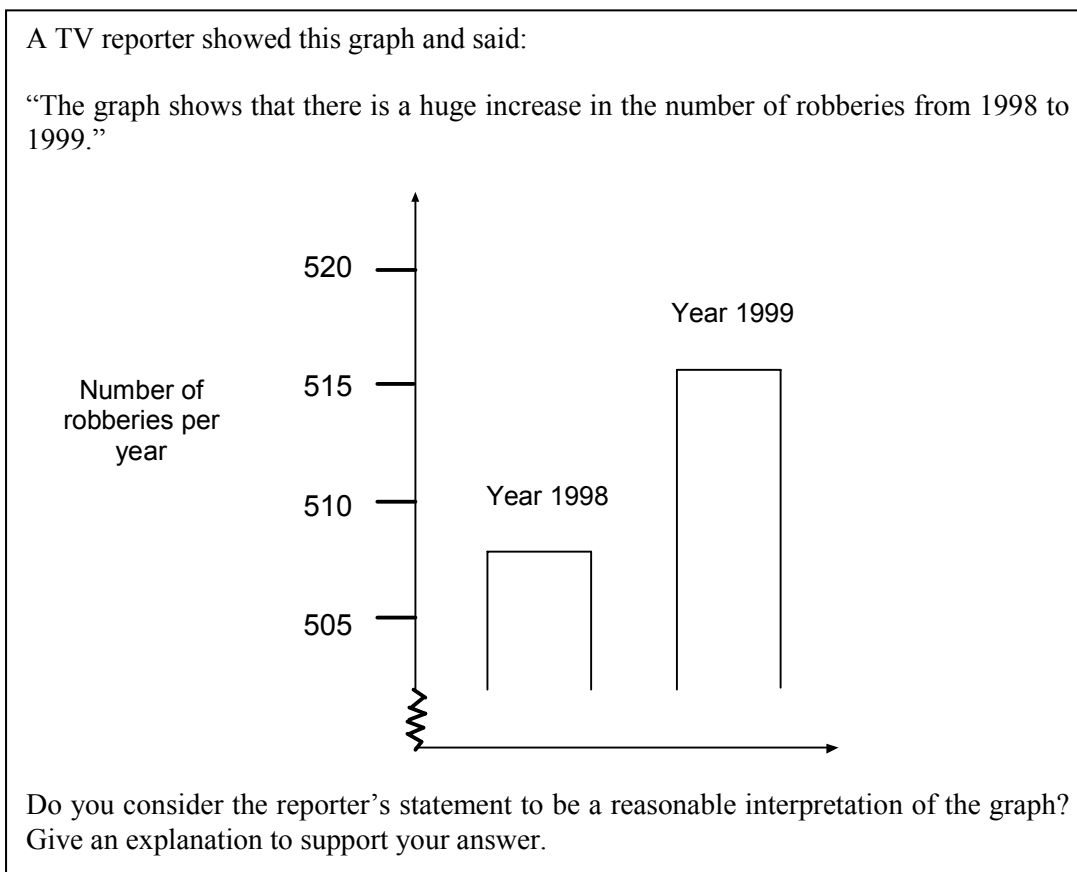
The value of  $b =$   zeds

The PISA item in Figure 6 is an example of an item that is explicitly interested in the ability of 15 year olds to interpret, use and apply some formal mathematical knowledge mainly related to school based curriculum. In PISA the context would be *Societal* with a content area of *Change and relationship*; and into the response type of *Employ*. It would not be expected to appear in PIAAC as it requires quite specific and technical understanding of school-based mathematics related to recognising and understanding the conventions and behaviour of linear equations. The item response type is *Constructed Response Auto-coded*. This is an example of a computer-based item with at least two types of interactivity: animation and automatic calculation. It also has a significant amount of text, and the language in the stimulus and question requires a high level of mathematical understanding, both of text and symbols.

The item shown in Figure 7 is an old PISA item that satisfies criteria for both assessments, and in fact similar items have been used in ALL and PIAAC. It fits within the contexts of *Societal or community* in PIAAC or *Societal* in PISA; the content areas of *Data and chance* or *Uncertainty and data* respectively;

and into the response types of *Interpret*, *evaluate* and *Interpreting, applying and evaluating* in PIAAC and PISA respectively. The item response type is *Constructed Response Expert*; and is an item that has no computer-based interactivity. As it is a *Constructed Response Expert* type of item it could **not** appear in the computer-based delivery of PIAAC unless it was considerably restructured. It has a small amount of text, but the language in the stimulus and question requires careful reading and interpretation.

Figure 7: PISA item



## CHAPTER 5: SUMMARY AND CONCLUSIONS

To date, the literatures on adult numeracy and school-based mathematical literacy have developed in some isolation, despite the conceptual linkages between numeracy and mathematical literacy and despite significant and quite related theoretical advancements in both domains. Likewise, the design of international assessments of numeracy and mathematical literacy has progressed along related lines but without much direct dialogue between the two development efforts<sup>10</sup>. The present paper is the first to analyse in detail the commonalities and differences between the conceptualisation and the implementation of the constructs of numeracy and mathematical literacy in two large scale international assessments conducted by OECD, i.e. PIAAC and PISA. Such an analysis is needed in light of the relevance of both studies to a wide range of policy makers and stakeholders (Breakspear, 2012) and to the understanding of the importance of mathematical competencies across the lifespan. Further, countries that have participated both in PISA and PIAAC may wish to better understand to what extent the results from the two assessments can be connected.

With the above in mind, this paper has been organized with four main chapters. Chapter 1 examined and compared the policy goals of PISA and PIAAC, and Chapter 2 focused on the conceptualisation of the constructs of mathematical literacy and numeracy, respectively. Chapter 3 compared the assessment (methodological) frameworks used in PISA and PIAAC, while Chapter 4 focused on the operationalisation of the constructs in the actual assessments. This concluding chapter summarises conceptual and assessment-related commonalities and differences, and presents some conclusions regarding the comparability of PISA and PIAAC results.

We emphasise at the outset that the separation between conceptual and assessment aspects is mainly aimed to assist clarity of presentation; the two topics should be viewed in integration, and indeed in this chapter they are sometimes discussed in combination. This is because a conceptual framework for an assessment interacts with the assessment (methodological) framework and both jointly impact the findings and their interpretation. Eventually, what the *results* of any assessment can tell us depends on the combination of the conceptual and assessment frameworks, and will always be shaped by the various technical decisions made and the constraints in the field (i.e. how PIAAC or PISA were implemented in terms of sample sizes, item pools, computer technology, etc).

### 5.1 Conceptual issues: commonalities and differences

Based on the detailed comparison of the two frameworks for PISA and PIAAC in Chapter 2 and of the assessment frameworks and item pools in Chapters 3 and 4, it is apparent that both PISA and PIAAC describe and cover similar territories. On the conceptual level, numeracy and mathematical literacy certainly are related constructs in terms of the core ideas that underlie them; this is reflected in several ways in the definitions of the constructs:

- Both constructs refer to the ability of individuals to cope with tasks that are likely to appear in the real world, and that contain mathematical or quantitative information or that require mathematical or statistical skills and knowhow.
- Both constructs focus on how well individuals can use their mathematical knowledge and skill to solve problems stemming from pragmatic (i.e. real-world) needs or demands, and to ‘engage’,

---

<sup>10</sup> As explained in Chapter 2, the design of the Numeracy assessment in PIAAC is based in large part on the conceptual frameworks and item pools developed for ALL. Work on this aspect of ALL started in 1998, separate from the efforts to design PISA 2000 which occurred roughly during the same timeframe.



manage, and understand various tasks in the world around them—rather than addressing decontextualised mathematical tasks.

- Both PISA and PIAAC describe mathematical literacy or numeracy as *not* synonymous with a minimal or a low level of mathematical knowledge and skills. That is, both assessments view the constructs as describing competencies lying on a continuum, i.e. individuals could be placed on a scale from low levels to high levels.

That said, there are aspects in the conceptualisation of mathematical literacy and numeracy that should be noted, in connection with the policy goals and ecology of each assessment. We divide them below into three separate but certainly related points.

First, the conceptual description of mathematical literacy in the PISA framework has shifted from PISA 2000 to PISA 2012 and became more saturated with direct references to the “mathematics” in mathematical literacy. Such an emphasis was already evident in the definition of mathematical literacy in PISA 2000 (“*Mathematical literacy is an individual’s capacity to identify and understand the role that mathematics plays in the world...*”) but became more pronounced in the language chosen for PISA 2012 (“*Mathematical literacy is an individual’s capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena. It assists individuals to recognize the role that mathematics plays in the world ...*”).

In contrast, PIAAC is focused from the outset on the tasks that adults have to cope with in the world, and hence the conceptualization of adult numeracy is phrased in a different way that emphasises the *tasks* and what is required to manage them effectively (“*Numeracy is the ability to access, use, interpret, and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life*”). Also, the conceptualization of numeracy in PIAAC includes a unique element designed to help the operationalization of ‘numeracy’ in an actual assessment: the description of a sub-construct termed *numerate behaviour* which “*involves managing a situation or solving a problem in a real context, by responding to mathematical content/information/ideas represented in multiple ways*”.

In our view, the difference in emphasis noted above stems from the trajectory of development and from the ecology (e.g. community of scholars and stakeholders, policy goals) in which each construct evolved, described in Chapters 1 and 2 and further clarified in the respective framework documents for PISA and PIAAC (see also OECD 2013a,b). It seems that mathematical literacy has been shaped in PISA based on a mix of perspectives, both internal (i.e. what is mathematics, what are the goals of mathematics education) and external (i.e. real-world manifestations of mathematics).

In the decade after the release of PISA 2003, pressures have mounted to link the construct of mathematical literacy and the assessment itself closer to school-based mathematics curricula, i.e. emphasize the ‘mathematics’ in mathematical literacy. This trend is reflected in the greater emphasis placed in PISA 2012 compared to PISA 2003 on describing the mathematical modelling (or mathematisation) cycle and other enabling processes that underlie how individuals cope with mathematical literacy tasks. In contrast, the references to mathematics in the conceptual framework describing adult numeracy in PIAAC are less visible, since the design of the PIAAC numeracy construct has been free from school-related considerations, and has aimed to reflect from the outset the actual external demands set on adults and the range of actual competencies that adults need to be able to cope with. For this reason, messages regarding mathematisation and mathematical modelling are not emphasized in the opening statements of the PIAAC numeracy framework – but are instead subsumed elsewhere in that framework, under titles such as “enabling processes”.

Second, the constructs and their implementation in the actual assessment are also shaped by the policy goals and assessment ecology. As explained in Chapter 1, PISA and PIAAC have different policy goals and these have several ramifications that are not necessarily obvious to those familiar only with school-based assessments or only with adult assessments, hence reiterated here:

- PISA aims to examine skills of *students* who are in school at present, and have typically been to school for 9 or 10 grades, but before they have had much exposure to the world of work or to the full range of tasks faced by adults. As a result, the conceptualisation of mathematical literacy, and what constitutes “low proficiency” in this regard, are informed in PISA by the criterion of what is expected from a school *graduate*. Further, the design of PISA tasks (i.e. assessment items) is somewhat constrained because school students have limited familiarity with the adult world.
- In contrast, PIAAC aims to examine the full range of skills in the *entire adult population* (in ages 16-65) across countries whose economic and social systems may differ in many ways. As a result, the design of the PIAAC methodology and assessment takes for granted that some adults in all countries have had relatively little or no formal education, or have been out of school and hence not familiar with formal symbols and school-based terminologies, or may have limited language skills, yet their proficiency also has to be described [on the same reporting scale]. Thus, the range of skills to be covered by the assessment instruments is wider in PIAAC compared to PISA, and PIAAC has to start its assessment at a lower point on the continuum. This reality affects the number and spread of items that can be included to cover any one level.
- In addition, the different policy goals led to quite different sets of variables being examined in the Background questionnaire in PIAAC and in the Student and School questionnaires that accompany PISA. PIAAC aims to assess not only antecedents and background factors that may be associated with the measured skills, but also economic and social outcomes (see Figure 1) and related correlates such as job characteristics. Hence, some of the testing time in PIAAC has to be allocated to such key variables that are not paramount in PISA, which in turn limits the amount of time that can be allocated to the cognitive assessments in PIAAC.

Third, the demarcation in PIAAC between numeracy and other skills has been informed by the adoption of a comprehensive view of adult competencies which has characterised ALL and prior studies of adult skills as well. It is taken for granted that when adults face *any* real-world task, *all competencies interact* and cannot be fully artificially separated in practice. These points affect both the design of assessment items (e.g. whether texts are seen as external to the competency being assessed or an inherent part of adult life) but also affect interpretation of the final results. As a consequence, results for cognitive domains measured in PIAAC, as in prior adult skills surveys, are sometimes reported *side by side*. In contrast, PISA findings pertaining to mathematical literacy are not connected with findings regarding other PISA domains, i.e. reading literacy or science literacy. This practice is in part affected by the PISA design, since in every PISA assessment cycle one domain receives expanded coverage compared to the others—but in our view it also reflects the independence of these domains in school curricula which leads to treating each competency domain as a stand-alone entity.

We believe that despite the differences noted above in the conceptual frameworks for numeracy and mathematical literacy, the actual distance between the conceptualisations of numeracy and mathematical literacy is not large, when the trajectories of development and intellectual ecologies as well as policy needs of each survey are taken into consideration. The general spirit of the definitions is quite similar as summarised above. While using somewhat different terminologies, the constructs certainly refer to quite similar building blocks, as evident in the analysis of the four content areas and the four contexts analysed in detail in Chapter 4. Yet, differences between the nature of what is being assessed in each program are more evident when we take into account methodological and implementation issues, some of which were already noted above, and others summarised below.

## 5.2 Assessment and methodology issues: commonalities and differences

When looking further, the PISA and PIAAC programs use methodologies that have on the surface many similarities in their approaches to assessing mathematical literacy and numeracy. For instance, both assessments employ a mix of paper-based and computer-based assessments. They present respondents with items reflecting real-world tasks using multiple types of representations (e.g. text, graphs, symbols), and reflecting four key content areas (using labels such as quantity & number; data & uncertainty; pattern & relationships; dimension & shape) with almost the same proportions. Further, both assessments employ IRT scaling techniques to estimate overall proficiency distributions, and report results in terms of 5-6 proficiency levels.

The commonalities noted above, together with the overall similarity in the constructs noted in the opening of section 5.1 above, could lead to the conclusion that it would appear quite valid to use the results and data from PISA's mathematical literacy and PIAAC's numeracy for comparative purposes. Yet, section 5.1 above already mentioned several differences between PISA and PIAAC in terms of policy goals, conceptualization, and assessment ecology, which do have ramifications for the assessment framework implemented in each program. Using these differences as a point of departure, we note that Chapters 3 and 4 presented a detailed analysis of commonalities and differences regarding the assessment frameworks that affect the nature of what is being assessed. Below we summarize and elaborate on selected key points in this regard.

### 5.2.1 *The number, characteristics, and range of assessment items*

As explained in Chapters 3 and 4, the planned time for testing adults in PIAAC is much shorter than the time available for the assessment of students in PISA, given the time constraints on assessment in a household survey compared to the assessment of students in schools; further, the planned testing time has to be allocated both to cognitive measures and background questionnaires, which differ in PISA and PIAAC. As a result, the amount of time allocated to the assessment of any cognitive domain in PIAAC, including numeracy, is short and this limits the number of items that can be given to any individual adult respondent. As noted in Chapter 3, the extended time allocated to assessment of mathematical literacy in PISA 2003 and PISA 2012 has allowed for the administration of a much larger number of mathematical literacy items per respondent (close to 3 times as much) compared to the number of numeracy items per respondent in PIAAC, which is more similar to the number of mathematical literacy items in PISA 2006 and 2009.

Further, the computer-based testing environment that could be employed in PIAAC for the assessment of most individuals was planned for immediate, automatic scoring of responses in order to allow the use of adaptive testing and routing processes necessary for efficient ability estimation. The need for automatic scoring prevented the use of open constructed response items for the majority of respondents in PIAAC and precluded the use of human scorers, limiting the range of possible item types (although a range of response options were available in the computer-based assessment in PIAAC that are not multiple-choice in type, as explained in section 4.4). That said, some open-ended (constructed-response) items *were* included in the paper-based assessment of PIAAC, but it is assumed this were mostly given to persons who are older or less comfortable with ICT, and these characteristics may also be correlated with lower SES in some cases. In contrast, and as explained in Chapter 4, PISA's main assessment relies on the use of printed booklets and on scoring by human scorers, and these features have allowed the inclusion of a wide range of item types in all cycles of PISA.

Beyond test length and item type issues, there are some differences between PISA and PIAAC in terms of item difficulty levels due to the differences in respondents' ages. PISA, especially in 2012, but also in earlier cycles, has had an interest in collecting evidence about the ability of 15 year-olds to use and

apply formal school curriculum based maths skills and knowledge. However, since PISA is administered only to 15-year olds who normally are in grade 9 or 10, it has not traditionally aimed to examine basic or simple mathematical skills<sup>11</sup>. In contrast, PIAAC must assess the full range of numeracy skills in the entire adult population, and hence includes items at a wider range of difficulties, including some aimed at much lower level mathematical skills.

The differences in orientation of PISA and PIAAC also affect their internal complexity schemes (see section 4.6). The two complexity schemes appear distinctively different in terms of the number of factors and their descriptions, although closer scrutiny reveals some significant commonalities. Hence, it would be of much interest to further examine whether there could be ways of integrating aspects of the two schemes to improve their use in predicting the difficulty levels of mathematical literacy or numeracy tasks, independently of the age of the respondents.

### 5.2.2 Comparability of reported results

As explained in detail in section 3.6, several general differences exist between PISA and PIAAC in terms of *measurement scales* (200-800 in PISA, 0-500 in PIAAC), *response probabilities* (RP of .62 in PISA versus .67 in PIAAC), and *proficiency levels* used for reporting purposes (six levels in PISA, five levels in PIAAC). The combined effect of these three differences, which are not specific to the assessment of mathematical literacy or numeracy per se, is that a simple comparison of reported results from PISA and PIAAC has to be done with much caution. In our view, it is not possible to directly compare the proportion of persons who are, for example, at “Level 2” (or any level) in PISA and in PIAAC across countries, because “Level 2” (or any level) on the two assessments does not mean being at the same absolute level of proficiency. This is first of all due to the differences both in the number of reporting levels and the underlying RPs – but also due to the different spread of items across difficulty levels noted in the prior section, and other differences in the nature of the items examined in Chapter 4 and reiterated in this chapter.

Beyond these issues, it is useful to reflect on the possibility of comparing performance of the age cohorts of people who participated as students in earlier cycles of PISA and later in PIAAC (where data collection took place in 2011-2012). This could be of possible interest to some countries in order to try and explain the performance of adults in terms of what they studied earlier in school, or to gauge the impact of a change in national educational policy on the proportion of people in different proficiency levels. However, such an analysis has to face the realities of the two assessment programs. A typical national PIAAC sample includes 5000 completed cases covering a range of about 49 years (from age 16 to 65). Assuming a flat age distribution for the sake of simplicity, about 100 respondents will be included in a single year age group, or about 200 cases for the two age groups which are being compared on a cross-sectional basis to the PISA cohort. However, out of these 200 PIAAC respondents, *less than two thirds* would have taken the numeracy assessment, given the PIAAC assessment and routing design (see Chapter 2). This means that the resulting subsample of PIAAC respondents whose performance in numeracy is of interest would be further reduced.

The above considerations imply that there is a limitation on the types of comparisons possible for PISA cohorts who later participated in PIAAC. Let us assume, for example, that researchers wish to compare mathematical literacy scores of those tested in PISA 2000 (at ages 15-16) to the numeracy scores of the comparable age cohort tested in PIAAC in 2011 (i.e. when this age cohort would be in ages 26-27).

---

<sup>11</sup> Indeed, few easy or simple items were included in PISA 2000 to 2009, necessitating the use of a “Below Level 1” reporting category. This has changed in PISA 2012, where relatively easy items were intentionally developed and added to the assessment in order to better describe the many students who are at the low end of the proficiency distribution.

We believe it is possible to compare, with caution, *average* (mean) levels of proficiency across PIAAC and PISA for such a subgroup and certainly for a national sample, in order to get a sense for the relative standing of countries in this regard. However, it may not be possible to go beyond a comparison of national means and compare *the distributions* of proficiency scores in mathematical literacy of those who took PISA at a certain year to the distribution of numeracy scores in PIAAC of the comparable two-year cohort in PIAAC. This is because, as explained earlier, the group of respondents in a specific age range (e.g. 26-27 only) who were tested in numeracy in PIAAC would be too small to allow for further breakdown in terms of the proportion of respondents who were found to be at each one of the 5 or 6 performance levels in numeracy as some of these subgroups would be too small for a credible statistical comparison.

If it was desired to conduct more detailed comparisons of PIAAC and PISA results, further research would need to be done. For example, a rating study could be designed to compare how experts perceive and rank the relative difficulty levels of the items from both assessments, using the existing complexity schemes described in Chapter 4. Additionally, an equating study with a sufficiently large sample could be designed to compare the actual difficulty levels of PISA and PIAAC items taken by the same group of individuals. Such and related studies have of course associated costs and limitations of their own, but are mentioned here because they could help to examine the alignment of numeracy and mathematical literacy items and scales in terms of relative difficulty levels, and enable better comparison of results from both assessments.

### 5.3 Conclusions

Both assessments of numeracy in PIAAC and mathematical literacy in PISA appear to have substantial conceptual similarities and quite a few practical commonalities in the nature of their test items and their design principles, as well as the range of content areas and skills they cover. The two surveys are highly consistent in their descriptions and structures for contexts and real world content classifications, along with how they describe the types and breadth of responses and actions expected of the respondents. Yet, results from PISA and PIAAC are shaped by the different characteristics and inevitable constraints of the methodology of implementation of each assessment in the field (e.g. test design, use of paper-based versus computer-based items as the main assessment tool, automated versus human-based scoring, etc).

Although there are several differences in the item types in the two assessments, eventually in PIAAC the data from the computer-based and paper-based portions of its assessment (see Chapter 2) is jointly used to generate population-level estimates of skills distributions, just like in PISA. This analytic approach compensates for the relative limitations imposed on the PIAAC computer-based assessment and has the potential to generate proficiency estimates equal in value to those in PISA. PISA 2012, with its more comprehensive range of item types and more interactive computer-based assessment, will enable richer and extended descriptions of sub-components of mathematical literacy compared to the information that can be generated by the numeracy assessment in PIAAC. However, PIAAC has aimed from the outset to innovate in several other ways in its overall design and collection of correlates, in order to maximise the analytic potential regarding policy-related questions pertaining to the contribution of skills to social and economic outcomes as well as many other issues.

As a result of the various methodological issues noted earlier, we believe that findings and proficiency distributions from both PISA and PIAAC should not be directly compared without much caution and additional research. Yet, we also believe that despite various differences noted in this paper, the two assessments are both covering a comprehensive and quite common set of variables related to the use and application of mathematical and statistical knowledge and skills in real world contexts. Both PISA and PIAAC can help policy makers, educators, researchers and other stakeholders to understand to what extent the knowledge and skills possessed by adults and young adults are synchronised with the nature of the tasks facing adults in modern societies and information-rich economies.

## REFERENCES

- Breakspear, S. (2012), "The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance", *OECD Education Working Papers*, No. 71, OECD Publishing.  
doi: [10.1787/5k9fdfqffr28-en](https://doi.org/10.1787/5k9fdfqffr28-en)
- Bynner, John & Parsons, Samantha (2005). *Does numeracy matter more?*, National Research and Development Centre for Adult Literacy and Numeracy (NRDC), London
- Coben, D., O'Donoghue, J., & FitzSimons, G. (Eds.) (2000), *Perspectives on adults learning mathematics*. London: Kluwer.
- Cockcroft, W. H. (Chair of the Committee of Inquiry into the Teaching of Mathematics in Schools). (1982), *Mathematics Counts*. London: HMSO. Crowther Report, UK (1959) 15-18: report of the Central Advisory Council of Education (England) vol. 1. London: HMSO.
- De Lange, J. (2003), Mathematics for Literacy. In B. L. Madison, & L. A. Steen (Eds.), *Quantitative literacy: Why numeracy matters for schools and colleges*. Proceedings of the National Forum on Quantitative Literacy held at the National Academy of Sciences in Washington, D.C. on December 1-2, 2001. National Council on Education and the Disciplines Princeton, New Jersey, 2003. Retrieved July 10, 2012 from: [www.maa.org/ql/qltoc.html](http://www.maa.org/ql/qltoc.html)
- Ferrini-Mundy, J. & Schmidt, W. H. (2005), International comparative studies in mathematics education: Opportunities for collaboration and challenges for researchers. *Journal for Research in Mathematics Education*, 36(3), 164-175.
- Gal, I. (1997), Numeracy: Imperatives of a forgotten goal. In L. A. Steen (Ed.), *Why numbers count: Quantitative literacy for tomorrow's America* (pp. 36-44). New York: The College Board.
- Gal, I. (2002), Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25.
- Gal, I., van Groenestijn, M., Manly, M., Schmitt, M. J., & Tout, D. (2005), Adult numeracy and its assessment in the ALL survey: A conceptual framework and pilot results. In Murray, S. T., Clermont, Y., & Binkley, M. (Eds), *Measuring adult literacy and life skills: New frameworks for assessment* (pp. 137-191). Ottawa: Statistics Canada.
- Ginsburg, A., Cooke, G., Leinwand, S., Noell, J., & Pollock, E. (2005), *Reassessing U.S. International mathematics performance: New findings from the 2003 TIMSS and PISA*. Washington, DC: The American Institutes for Research. Eric Document 491624). Retrieved Feb 1, 2012, from <http://eric.ed.gov/PDFS/ED491624.pdf>
- Ginsburg, L., Manly, M., & Schmitt, M. J. (December 2006), *The components of numeracy*. (NCSALL Occasional Paper). Harvard Graduate School of Education, Cambridge, MA: National Center for the Study of Adult Learning and Literacy. Online: [www.ncsall.net/](http://www.ncsall.net/)
- Hagedorn, L., Newlands, J., Blayney, I., & Bowles, A. (2003), *Frameworks for adult numeracy education: A survey and discussion*. Ontario, Canada: National Literacy Secretariat. Online: <http://www.nald.ca/fulltext/framework/framework.pdf>

- Hoyles, C., Wolf, A., Molyneux-Hodgson, S. and Kent, P. (2002), *Mathematical skills in the workplace*. London: Science, Technology and Mathematics Council. Online: <http://eprints.ioe.ac.uk/1565/1/Hoyles2002MathematicalSkills.pdf>
- Hudson, C. (2007), *Maths4Life - working mathematics for the 21st century?* National Research and Development Centre, NRDC, London. Online: [www.nrdc.org.uk/publications\\_details.asp?ID=119#](http://www.nrdc.org.uk/publications_details.asp?ID=119#)
- Jablonka, E. (2003), Mathematical Literacy. In A. J. Bishop, M. A. Clements, C. Keitel, J., Kilpatrick & F. K. S. Leung (Eds.), *Second international handbook of mathematics education* (pp. 75-102). Netherlands, Dordecht: Kluwer.
- Johnston, B. (1994, Summer), Critical numeracy? *Fine print*, Vol.16 (4), 7-12.
- Kilpatrick, J. (2001), Understanding mathematical literacy: The contribution of research. *Educational studies in mathematics* 47(1), 101–116.
- Kirsch, I. S., Jungblut, A., & Mosenthal, P. B (1998), The measurement of adult Literacy. In S. T. Murray, I. S. Kirsch, & L. B. Jenkins (Eds.), *Adult literacy in OECD countries: Technical report on the first International Adult Literacy Survey* (pp. 105-134). Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Lindquist, M. M. (June 2009), A Critical Evaluation of PISA's Measurement of Mathematics. Paper presented at the NCES Conference on the Program for International Student Assessment: What we can learn from PISA (June 2009, Washington DC). Retrieved June 2012 from: <https://edsurveys.rti.org/PISA>
- Madison, B. L. & Steen, L. A. (Eds.)(2003), *Quantitative literacy: Why numeracy matters for schools and colleges*. Proceedings of the National Forum on Quantitative Literacy held at the National Academy of Sciences in Washington, D.C. on December 1-2, 2001. National Council on Education and the Disciplines Princeton, New Jersey, 2003. Retrieved July 10, 2012 from: [www.maa.org/ql/qltoc.html](http://www.maa.org/ql/qltoc.html)
- Mullis, I. V.S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1998), *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study (TIMSS)*. Boston: Center for the Study of Testing, Evaluation, and Educational Policy.
- Murphy, S. (2010), The Pull of PISA: Uncertainty, Influence, and Ignorance. *InterAmerican Journal of Education for Democracy*, 3(10), 27-44. Online: [www.ried-ijed.org](http://www.ried-ijed.org)
- National Council of Teachers of Mathematics (1989), *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- Neidorf, T. S., Binkley, M., Gattis, K., & Nohara, D. (2006), *Comparing mathematics content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Assessments*. Technical Report. NCES 2006-029. Washington, D.C. National Center for Education Statistics. Online: <http://nces.ed.gov/pubsearch>
- Niss, M. (1996), Goals of mathematics teaching. In A.J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (Eds.), *International handbook of mathematics education* (11-48). Netherlands, Dordecht: Kluwer.

- OECD (1999), *Measuring Student Knowledge and Skills: A New Framework for Assessment*, OECD Publishing. doi: [10.1787/9789264173125-en](https://doi.org/10.1787/9789264173125-en)
- OECD (2002), "Definition and Selection of Competencies: Theoretical and Conceptual Foundations" - Strategy Paper, [Report DEELSA/ED/CERI/CD(2002)9], Paris.
- OECD (2004), *Learning for Tomorrow's World: First Results from PISA 2003*, PISA, OECD Publishing. doi: [10.1787/9789264006416-en](https://doi.org/10.1787/9789264006416-en)
- OECD (2010), *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*, PISA, OECD Publishing. doi: [10.1787/9789264062658-en](https://doi.org/10.1787/9789264062658-en)
- OECD (2010), Draft PISA 2012 mathematics framework. [Report EDU/PISA/GB(2010)2]. Paris: Author.
- OECD (2011), "Proficiency levels in PIAAC", [COM/DELSA/EDU/PIAAC(2011)14], Paris.
- OECD (2012), *Literacy, Numeracy and Problem Solving in Technology-Rich Environments: Framework for the OECD Survey of Adult Skills*. OECD Publishing. doi: [10.1787/9789264128859-en](https://doi.org/10.1787/9789264128859-en)
- OECD (2013a), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. OECD Publishing. doi: [10.1787/9789264204256-en](https://doi.org/10.1787/9789264204256-en)
- OECD (2013b), *The Survey of Adult Skills: Reader's Companion*. OECD Publishing. doi: [10.1787/9789264204027-en](https://doi.org/10.1787/9789264204027-en)
- OECD/Statistics Canada (2000), *Literacy in the Information Age: Final Report of the International Adult Literacy Survey*, OECD Publishing. doi: [10.1787/9789264181762-en](https://doi.org/10.1787/9789264181762-en)
- OECD /Statistics Canada (2011), *Literacy for life: Further results from the Adult Literacy and Lifeskills Survey*. OECD Publishing.
- PIAAC Numeracy Expert Group (2009), "PIAAC Numeracy: A Conceptual Framework", *OECD Education Working Papers*, No. 35, OECD Publishing. doi: [10.1787/220337421165](https://doi.org/10.1787/220337421165)
- PISA Mathematical Literacy Expert Group (2010), *PISA 2012 mathematics framework (draft)*. Retrieved August 20, 2011 from: [www.oecd.org/pisa/pisaproducts/46961598.pdf](http://www.oecd.org/pisa/pisaproducts/46961598.pdf)
- Romberg, T. (October 2001), Mathematical literacy: what does it mean for school mathematics. *Wisconsin School News*, 5-8. Retrieved July 15, 2012 from <http://hub.mspnet.org/index.cfm/9386>
- Schoenfeld, A. H. (2004), The math wars. *Educational policy*, 18(1), 253-286. doi: [10.1177/0895904803260042](https://doi.org/10.1177/0895904803260042)
- Secretary's Commission on Achieving Necessary Skills (1991), *What work requires of schools: A SCANS Report for America 2000*. Washington, DC: U.S. Dept. of Labor.
- Stacey, K., (2012), *The International Assessment of Mathematical Literacy: PISA 2012 Framework and Items*. Paper presented at the 12th International Congress on Mathematical Education, July 2012, Seoul, Korea.



- Statistics Canada and OECD (1996), *Literacy, economy, and society: First results from the International Adult Literacy Survey*. Ottawa: Author.
- Statistics Canada and OECD (1997), *Literacy skills for the knowledge society: Further results from the International Adult Literacy Survey*. Ottawa: Author.
- Statistics Canada and OECD (2005), *Learning a Living: First results of the Adult Literacy and Lifeskills Survey*. OECD Publishing. doi: [10.1787/9789264010390-en](https://doi.org/10.1787/9789264010390-en)
- Steen, L. A. (Ed)(1990), *On the shoulders of giants: New approaches to numeracy*. Washington, D.C.: National Research Council.
- Steen, L. A. (2001), *Mathematics and democracy: The case for quantitative literacy*. Washington, DC: Woodrow Wilson National Fellowship Foundation.
- Straesser, R. (July 2003), Mathematics at work: Adults and artefacts. In J. Maasz & W. Schloeglmann (Eds.), *Learning mathematics to live and work in our world: Proceedings of the 10th international conference on adults learning mathematics* (pp. 30-37). Linz, Austria: Johannes Kepler Universitat Linz.
- Tout, D, (2000), What is Numeracy? What is Mathematics? Paper presented at the International Congress on Mathematics Education 2000, Tokyo, Japan, August 2000.
- Tout, D, & Schmitt, M. J., (2002), *The inclusion of numeracy in adult basic education*. In J. Comings, B. Garner and C. Smith. (Eds.), *Annual review of adult learning and literacy: Volume 3*. San Francisco: Jossey-Bass.
- Turner, R. (2012), *Mathematical literacy – are we there yet?* Paper presented at Topic Study Group 6 (Mathematical Literacy), 12th International Congress on Mathematical Education, Seoul, Korea.
- Wickert, R, & Kevin, M. (1995), *No single measure: the final report*. Canberra, Australia: Department of Employment, Education and Training.
- Wu, M. (2010), "Comparing the Similarities and Differences of PISA 2003 and TIMSS". *OECD Education Working Papers*, No. 32. OECD Publishing. doi: [10.1787/5km4psnm13nx-en](https://doi.org/10.1787/5km4psnm13nx-en)

## RECENT OECD PUBLICATIONS OF RELEVANCE TO THIS WORKING PAPER

OECD (2012), *Literacy, Numeracy and Problem Solving in Technology-Rich Environments: Framework for the OECD Survey of Adult Skills*, OECD Publishing. doi: [10.1787/9789264128859-en](https://doi.org/10.1787/9789264128859-en)

OECD (2013), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing. doi: [10.1787/9789264204256-en](https://doi.org/10.1787/9789264204256-en)

OECD (2013), *The Survey of Adult Skills: Reader's Companion*, OECD Publishing. doi: [10.1787/9789264204027-en](https://doi.org/10.1787/9789264204027-en)

## THE OECD EDUCATION WORKING PAPERS SERIES ON LINE

The OECD Education Working Papers Series may be found at:

- The OECD Directorate for Education website: [www.oecd.org/edu/workingpapers](http://www.oecd.org/edu/workingpapers)
- Online OECD-ilibrary: [http://www.oecd-ilibrary.org/education/oecd-education-working-papers\\_19939019](http://www.oecd-ilibrary.org/education/oecd-education-working-papers_19939019)
- The Research Papers in Economics (RePEc) website: [www.repec.org](http://www.repec.org)

If you wish to be informed about the release of new OECD Education working papers, please:

- Go to [www.oecd.org](http://www.oecd.org)
- Click on “My OECD”
- Sign up and create an account with “My OECD”
- Select “Education” as one of your favourite themes
- Choose “OECD Education Working Papers” as one of the newsletters you would like to receive

For further information on the OECD Education Working Papers Series, please write to: [edu.contact@oecd.org](mailto:edu.contact@oecd.org).